RESEARCH ONLINE

University of Wollongong
Research Online

Australian Health Services Research Institute

Faculty of Business

2011

Measurement of disease in populations and risk assessment

Janet E. Sansoni
University of Wollongong, jans@uow.edu.au

Publication Details
J. E. Sansoni "Measurement of disease in populations and risk assessment", University of Canberra, 22 February 2011, (2011)

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

RESEARCH ONLINE

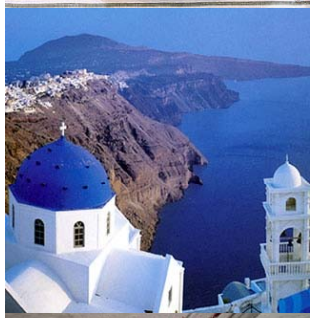
Measurement of disease in populations and risk assessment

Abstract
Powerpoint presentation presented at the University of Canberra

Keywords
assessment, disease, populations, risk, measurement

Publication Details
J. E. Sansoni "Measurement of disease in populations and risk assessment", University of Canberra, 22 February 2011, (2011)

This conference paper is available at Research Online: http://ro.uow.edu.au/ahsri/115



Measurement of Disease in Populations and Risk Assessment

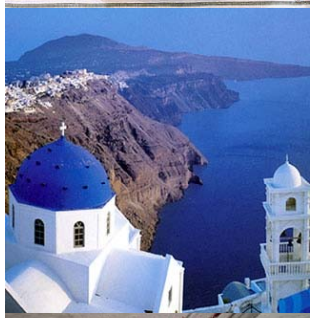
A/Prof Janet Sansoni

Centre for Research and Action in Public Health

Faculty of Health

University of Canberra

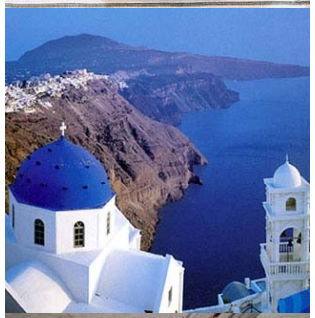
Acknowledgements: Gabrielle O'Kane, Catherine Itsiopoulos



Objectives of Lecture

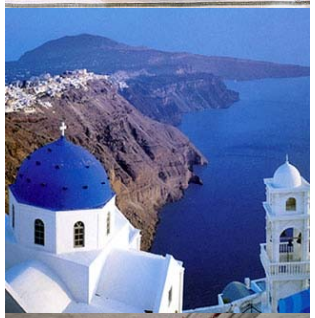
At the end of this lecture students will be able to:

- **Outline the reasons for measuring disease frequency**
- **Describe the difference between frequency counts (raw numbers), ratios, proportion and rates in measurement of disease**
- **Define prevalence and incidence (cumulative incidence and incidence rate)**
- **Calculate prevalence and incidence rates**
- **Write an equation to express the relationship between prevalence and incidence**
- **Explain why incidence data is more useful than prevalence data in considering aetiology of disease**
- **Discuss the benefit in using specific disease rates and age-standardised disease rates instead of crude disease rates**



Why measure disease frequency?

- **Measuring the amount of a disease in a community is essential to allow planning of health services**
  - **Building hospitals or community health centres, developing screening programs etc**
- **Measuring a health-related state, such as alcohol intake, saturated fat intake, physical activity levels in a population may help planning for future health needs and is also useful for the evaluation of public health programs and initiatives. Can you think of an example?**
- **Measurement of frequency of disease helps to assess the extent to which new risk factors or preventive strategies influence patterns of disease in the population. Can you think of an example?**
- **Measurement of patterns or the distribution of disease can identify disease clusters in the community where further action/intervention and investigation may be warranted**

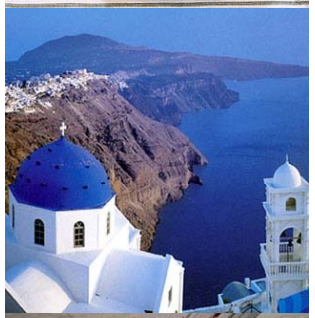


Count & Ratios

- **Count (raw numbers):** No. of cases of disease
  - There are flaws in using this type of crude measure but it played a useful role in the eradication of small pox (refer Webb et al.2005, P21).
  - Eg. A media article may say, “more people are dying of cancer than ever before.”
  - Why might this change be occurring?
  - See next slide on no. cases IHD by gender – can these figures be compared?
  - Also need to have a clear definition of the disease concerned (refer Fig 2.1 P29)

- **Ratio: a fraction (one number divided by another)**
  - **Eg. Number of beers drunk in one year divided by number of people in population (per capita) – Refer Box 2.1 Americans drink the most beer in the world (in total litres) whereas Czechoslovakia ranks 14th on this but on a per capita basis Czech’s drink the most beer whereas the Americans are now ranked 12th**

Ref: (see Box2.1; 2.7 Webb et al. 2005)



Differences in cancer mortality between males and females

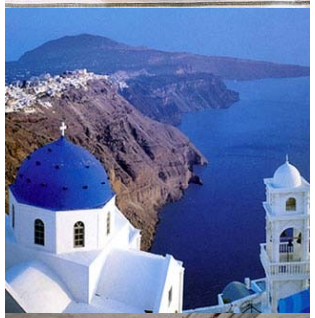
Table 2.3: Leading causes of death by number of deaths and sex, 1998

Males			Females		
	Cause of death	No.		Cause of death	No.
1	Ischaemic heart disease (410–414)	15,021	1	Ischaemic heart disease (410–414)	12,801
2	Lung cancer (162)	4,821	2	Cerebrovascular disease (430–438)	7,170
3	Cerebrovascular disease (430–438)	4,812	3	Dementia and related disorders (290, 294.1, 331)[(a)]	2,579
4	Chronic obstructive pulmonary disease (416.0, .8, .9, 490–492, 495–496)	3,325	4	Breast cancer (174)	2,542
5	Colorectal cancer (153, 154)	2,579	5	Colorectal cancer (153, 154)	2,165
6	Prostate cancer (185)	2,530	6	Lung cancer (162)	2,053
7	Suicide (950–959)	2,150	7	Chronic obstructive pulmonary disease (416.0, .8, .9, 490–492, 495–496)	2,026
8	Cancer of lymphatic tissue (200–208)	1,870	8	Cancer of lymphatic tissue (200–208)	1,600
9	Diabetes (250)	1,424	9	Diabetes (250)	1,327
10	Dementia and related disorders (290, 294.1, 331)[(a)]	1,294	10	Pneumonia (480–486)[(a)]	937

(a) Data are for 1996, as data for 1997 and 1998 are not comparable with previous years following the introduction of automatic coding by the Australian Bureau of Statistics in 1997.

Note: Codes refer to the International Classification of Diseases, 9th revision (ICD-9).

Source: AIHW National Mortality Database.



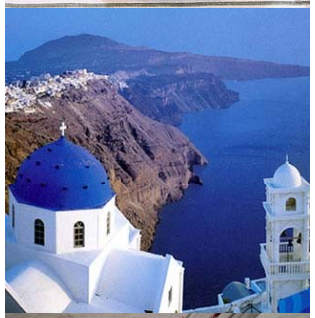
Proportions and Rates

- **Proportion:**
  - This is a particular type of ratio, as all those (people or things) in the numerator are also counted in the denominator. The number of people who develop the disease is divided by the total number of people in the population (including those with, and without, the disease)
    - Eg. Cases of diabetes/population at risk of getting diabetes – appears as a percentage or as a decimal – 8.4% Australian population have type 2 diabetes
- **Rate:**
  - This is like a proportion, but should contain some measure of time.
  - Eg. 2500 cases asthma/100 000 *per year*
  - *Case fatality ratio is a proportion and not actually a rate because it does not contain units of time*

Ref: (see Box 2.7 Webb et al. 2005)

Prevalence

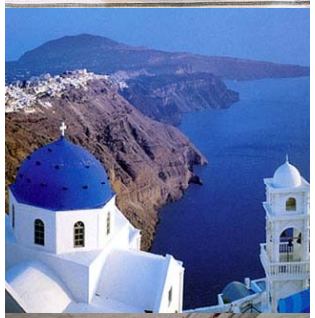
- Prevalence: This measures the proportion of the population with a disease (or health related event) at a given time.



- Point Prevalence: Number of persons with the disease in a population at a point of time/Number of persons in that population at that point of time

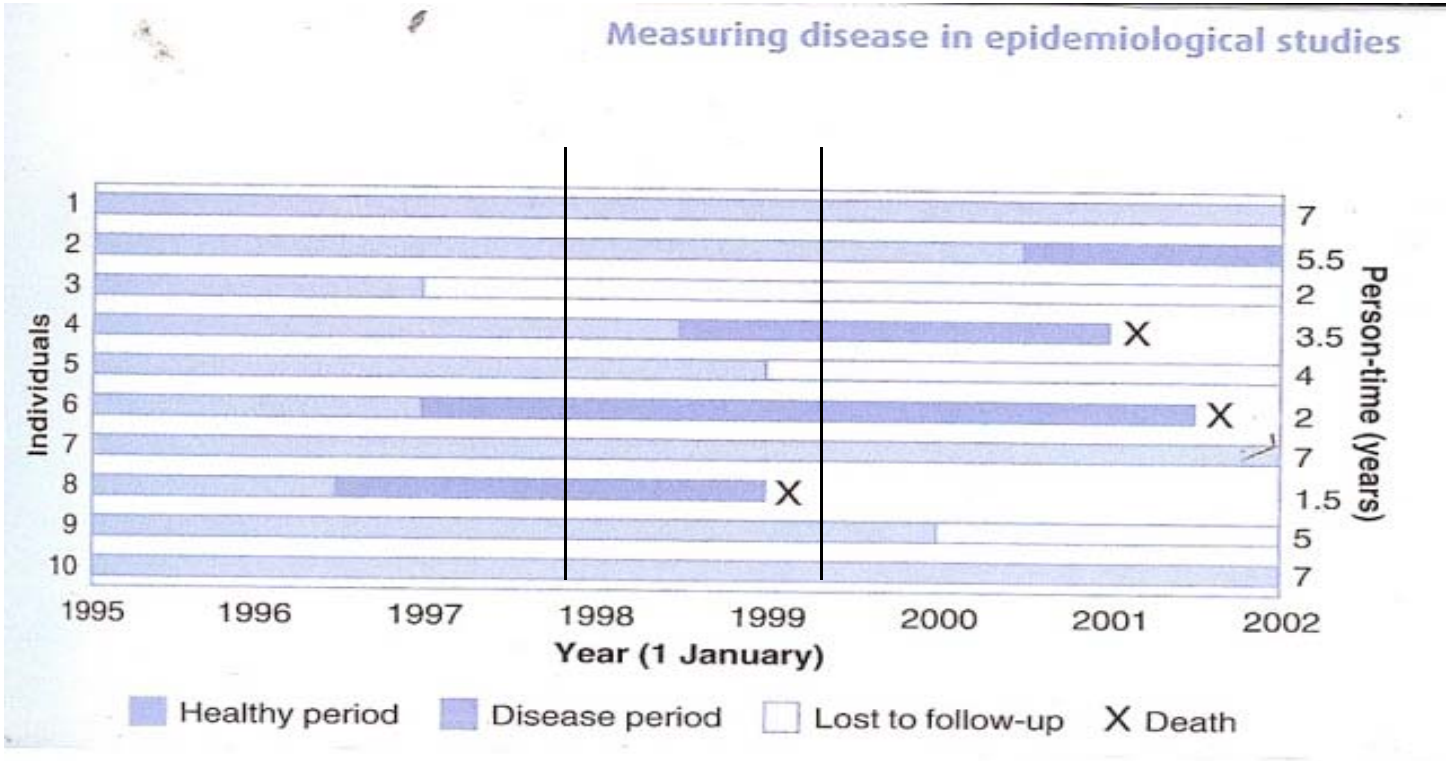
- Period Prevalence: Number of persons with the disease in a population during a period of time/Number of persons in that population at the midpoint of the period.
- See p. 38 & 39 of Webb 2005 for worked example (Next slide)

Ref: UNSW Centre for Public Health, 2002

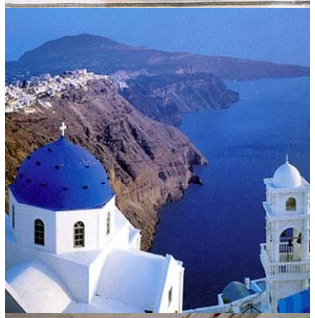


Period Prevalence

From Webb et al. (2005)



Estimate prevalence for 1 January 1988 and 30 June 1999



Examples

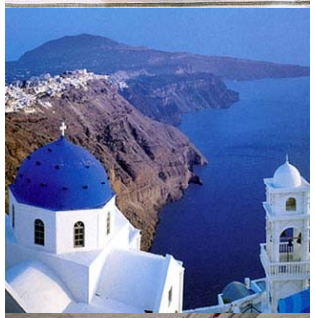
- Prevalence:
  - 1 Jan 1998, 2 people were sick out of 9 people left in our group on that date (one was lost to follow-up)
  - Prevalence = no. people with disease at given point in time/ total no. people in population

    = 2 / 9 = .22 or 22%

What was prevalence on 30 June 1999?

On 30 June 1999 2 people sick but only 7 people left in group (one died and two lost to follow-up)

Prevalence = 2/7 = .29 or 29%



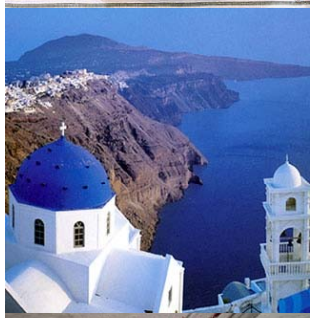
Incidence

- Incidence **refers to the number of *new* cases of disease occurring during a *defined period* in a *defined population*.**

- **There are two ways to describe** incidence
  1. **Cumulative incidence (risk):** the number of people in a defined population who develop a disease in a specified time period as a proportion of the population at risk at the start of the time period - **this is a proportion.**

  2. **Incidence rate (incidence density rate/person-time incidence rate/force of morbidity): –** the number of new cases of a disease in a defined population divided by the sum of the time periods experienced by all individuals in the population (total person time at risk) – **this is a rate, but risk and rate are used interchangeably.**

When estimating incidence rates from routinely collected stats on large populations, use the population at the midpoint of the period of interest for incidence rate – see p. 34 of Webb et al. 2005.

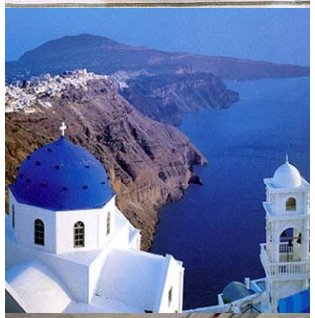


Attack Rate

- **Attack rate :**
  - alternative to cumulative incidence, used when the nature of the problem is of short duration (e.g. food-borne illness)
  - Attack Rate = number of new cases of disease during a period/population at risk at beginning of period
- **Application**
  - acute infectious disease episodes/epidemics or where the risk is short term

Incidence rate

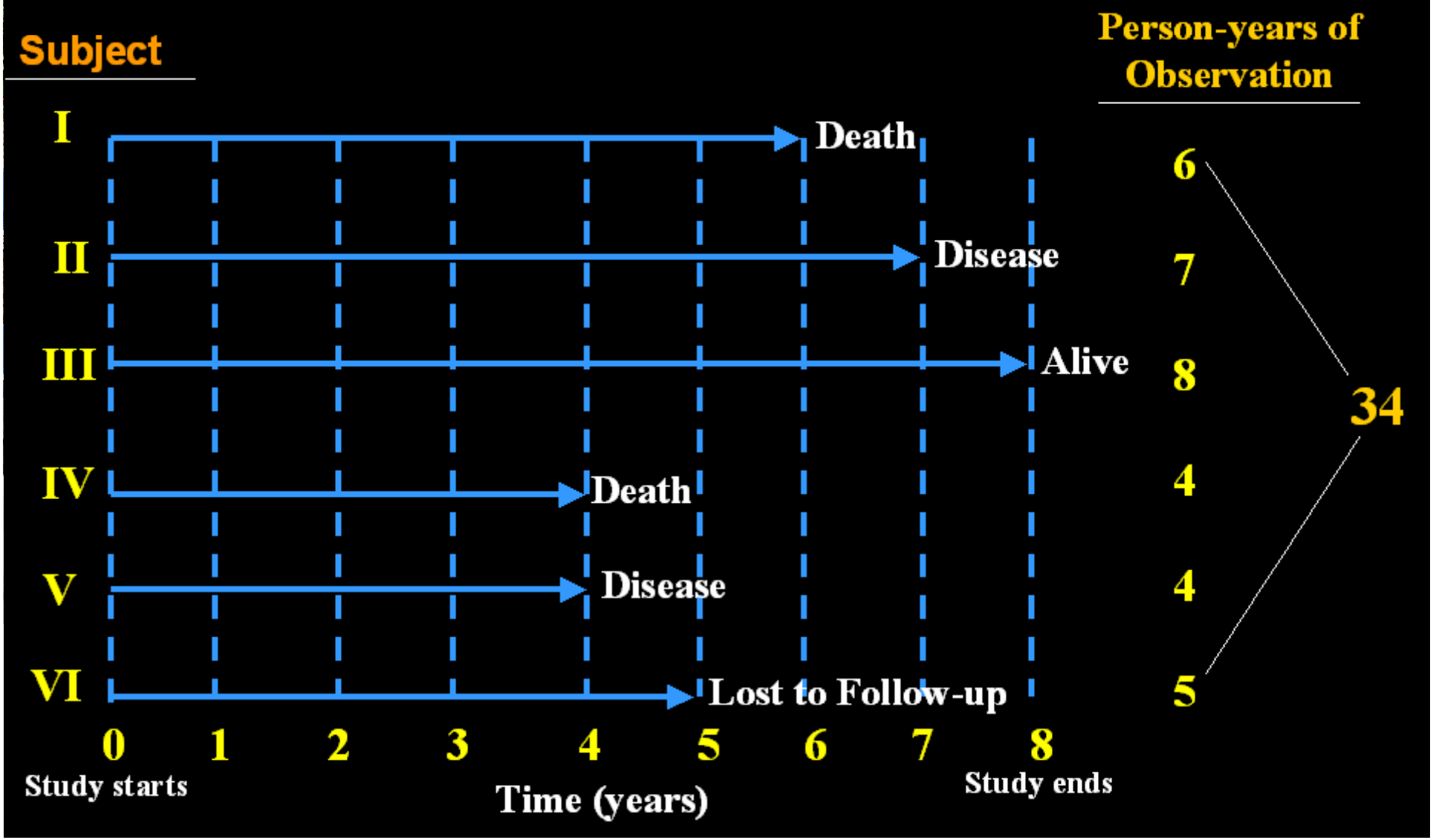
$$IR = \frac{\text{No. of new cases of disease during period}}{\text{Total person time at risk}}$$

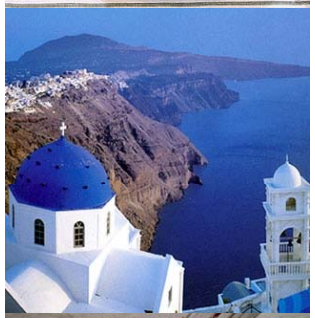


NB. In dynamic populations where people move in and out of populations with migration, lost to follow up or die, total person time at risk is more realistic.

- See next slide for example of person years from cohort study ->

Accumulation of person-years observed in an eight-year cohort study of six subjects





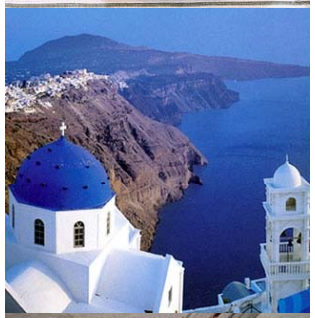
Prevalence and Incidence

P = existing cases or health events

→ magnitude

I = new cases or health events

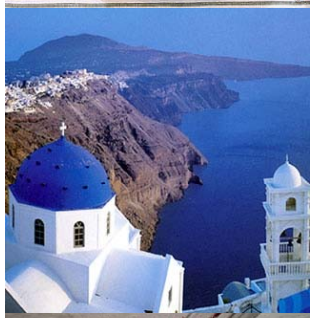
→ change



Prevalence and Incidence

- **Prevalence data are useful in determining the *extent* of a disease in the community**

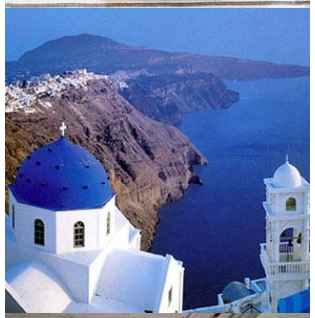
- **Incidence data is more closely linked to the aetiology or the cause of a disease. Why?**



The relationship between incidence and prevalence

- If 2 diseases have the same incidence but one lasts 3x longer then at any point in time you are more likely to find people suffering from the longer lasting disease

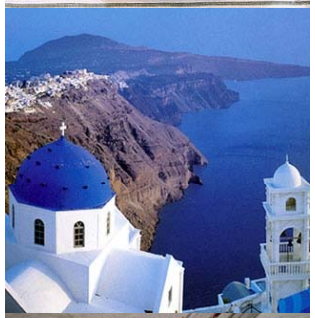
- Prevalence (P) is approximately equal to (~) the incidence (I) multiplied by the average duration of the disease (D)



What is the association between prevalence (P) and incidence (I)?

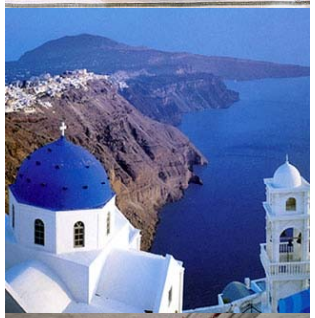
- **P ~ I x D (when prevalence is low)**
  - **Where D = duration of the disease**
    - **In a steady state population the numbers of people with disease and without disease are stable, so the number of persons added to the prevalence pool by new disease is equal to the number of persons removed from the disease by death or recovery.**
    - **For this formula the prevalence should be <~10%**

- **But where disease prevalence is high then the association differs**
  P/1-P ~ I x D
  - **Where 1-P include those who do no have the disease**
  - **The number of people with the disease divided by the number of people without the disease is approx equal to incidence x the average duration of the disease**



Example

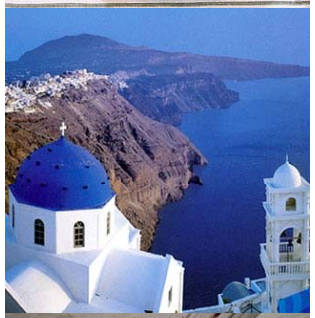
- Hep A has a relatively high incidence but because it is an infection of short duration the prevalence at any one point in time would be low
- Hep C has a much lower incidence but you have it for life and thus Hep C prevalence is much higher
- If a new treatment were developed for a disease what effect would this have on incidence and prevalence? (refer P 35, Webb et al. 2005)



Practice Example

- What % of children in kindergarten had a cold on their first week of school? (total class=100, no. of kids with colds=10 e.g. 10/100)
  - Is this prevalence, incidence, attack rate????
- What % of children who didn't have a cold in the first week, developed one during their next week?
  - Lets say 30 students got a cold in the second week
    - Calculate the cumulative incidence

Practice Example



Children with cold

- Attack Rate= number of new cases of disease during a period/population at risk at beginning of period
- (total class=100, no. of kids with colds=10) = 10%

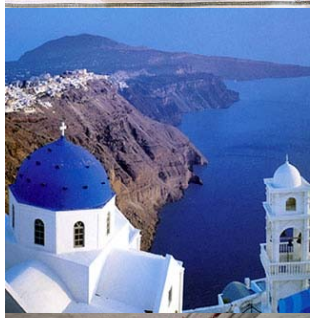
Cumulative Incidence:

CI= number of people who develop disease in specified period / number of people at risk of developing disease at start of period

Q. What % of children who didn't have a cold in the first week, developed one during their next week?

– Lets say 30 students got a cold in the second week

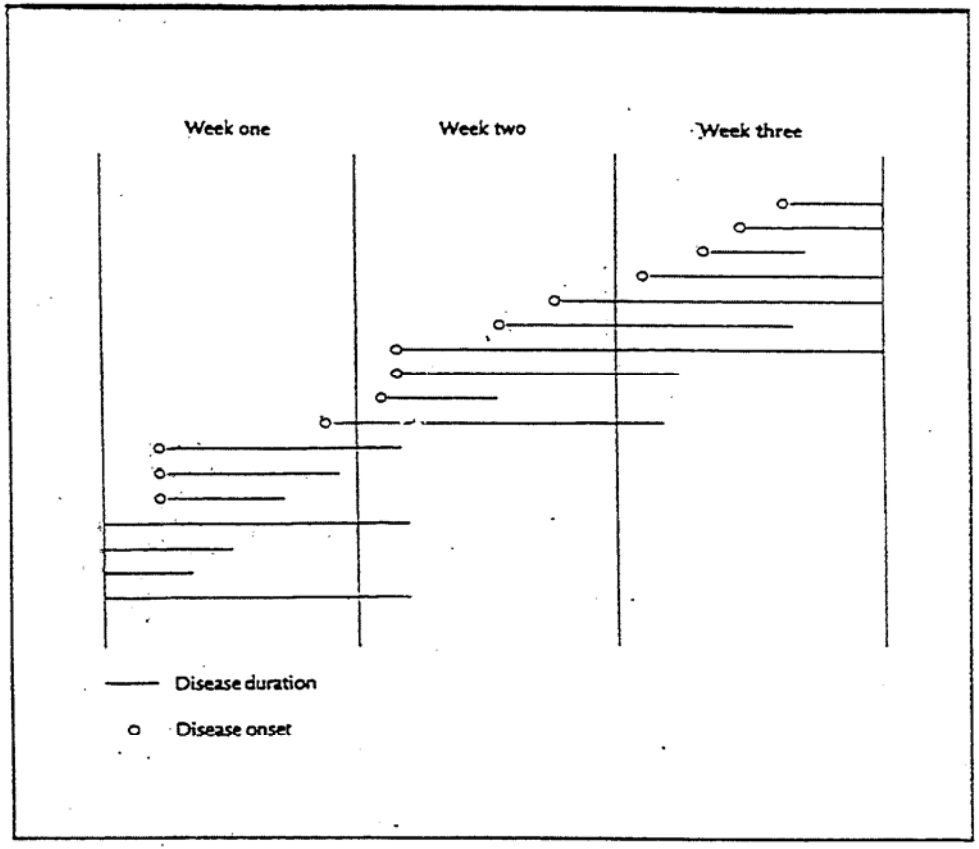
= 30/90 = 33%

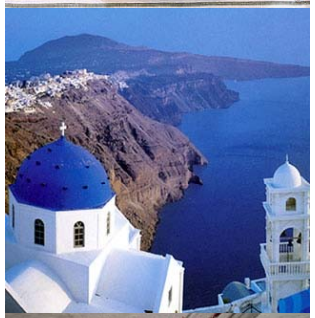


Practice example for prevalence and incidence measures

(ref: UNSW Centre for Public Health 2002)

Figure 1.2.2 Occurrence of disease X In 50 masters students over the first three weeks of semester





Questions

Assume there are no additions to the class and no ‘dropouts’.

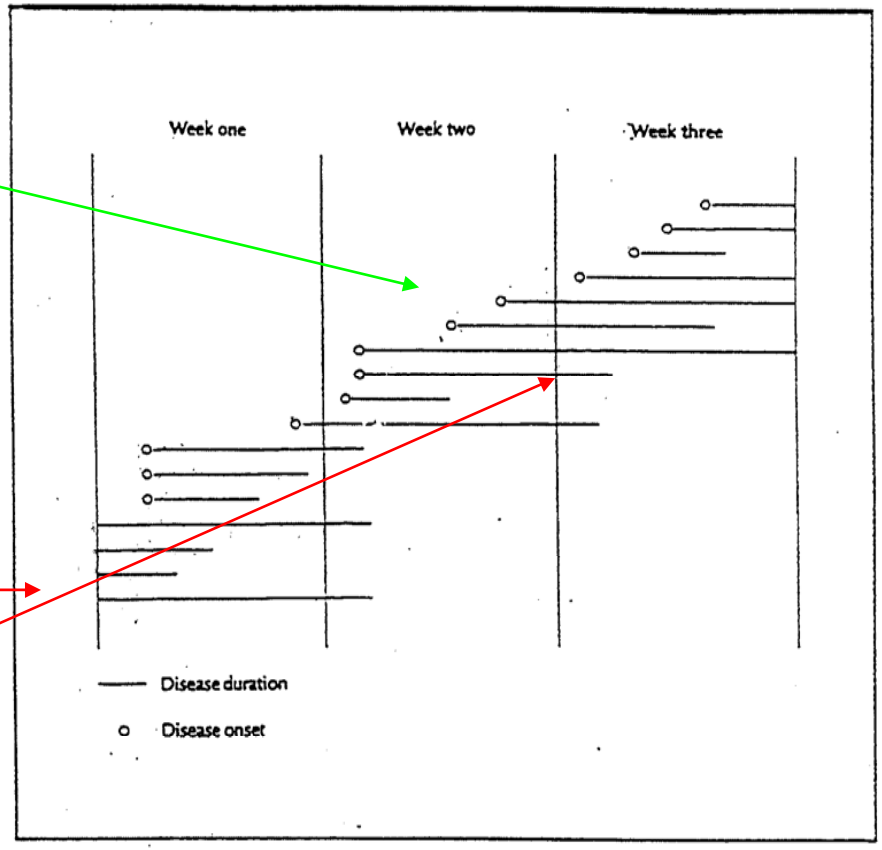
1. **What is the prevalence of disease X?**
   a. **At the beginning of week one?**
   b. **At the beginning of week three?**
2. **What is the cumulative incidence of disease X for week two?**
3. **What information do you require to calculate the incidence rate over the three week period? Calculate the incidence rate if possible**

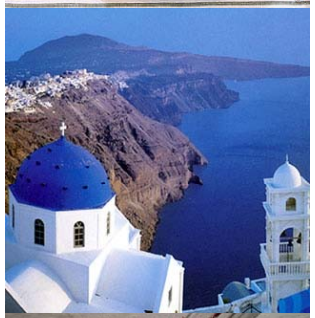
Practice example for prevalence and incidence measures

Cumulative Incidence for Week 2 =5/42 as 8 people already have disease (5 new cases/50-8 previously exposed cases)

Prevalence - 4 existing cases at beginning of week 1; 5 existing cases at beginning Week 3

Figure 1.2.2 Occurrence of disease X In 50 masters students over the first three weeks of semester





Incidence Rates vs. Cumulative Incidence

- Incidence Rate is equivalent to the average speed of a car at a particular point in time…the speed at which new cases are developing
- The Cumulative Incidence is like the distance travelled by the car during a specific period or time interval e.g. 60km travelled in I hour – it involves duration – new cases during the time interval

Factors Influencing Observed Disease Prevalence

Longer duration of the disease

Prolongation of life of patients without care

Increase in new cases (increased incidence)

In-migration of cases

Out-migration of noncases

In-migration of susceptible people

Improved diagnostic facilities (better reporting)

Shorter duration of the disease

Higher case-fatality rate of disease

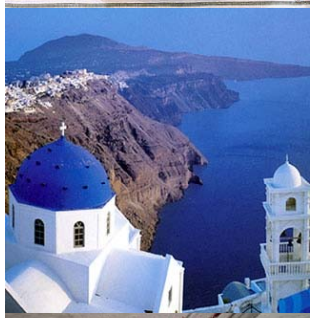
Decrease in new cases (decreased incidence)

In-migration of noncases

Out-migration of cases

Out-migration of susceptible people

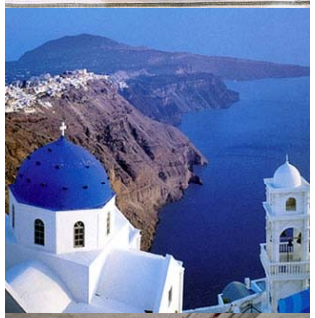
Improved cure rate of cases



Application of Incidence Data (rates)

Rates can be broken into 3 categories:

- Crude rates
- Specific rates (e.g. gender/age specific)
- Standardised (adjusted rates)
  - Relative measure
  - Adjust sample/ study data rates and measures according to distribution of the demographics of the total population from which the sample is derived
  - Why bother standardising the data?
    - Remove bias or the effects of different distribution for age, sex etc in total population



Crude Death/Disease Rate

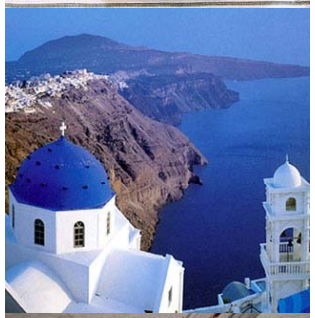
- **Crude rate of disease = Frequency of disease in a specified time period/total population**
- **Crude death rate = Number of deaths in a given year/reference population (at midpoint of year).**

Units used:

deaths/1 000; deaths/100 000; deaths/1 000 000

Why is the integer used for ‘death’ or disease important?

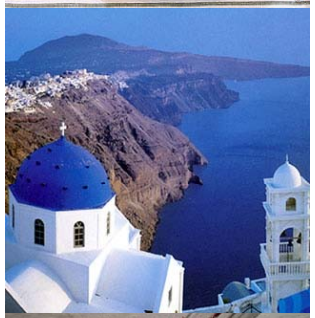
4 per 100 000 not 0.04 per 1 000



Commonly used Rates

- **Birth Rate = No. live births per year/population expressed per 1000**
- **Infant Mortality Rate (IMR) = No. infant deaths among infants aged 0-365 days/number of live births during the year**
- **Fertility rates= No. live births within a year /number of women aged 15-44 yrs at midpoint of the year**

  (see table 2.7 Webb – p. 52 –measures of mortality)

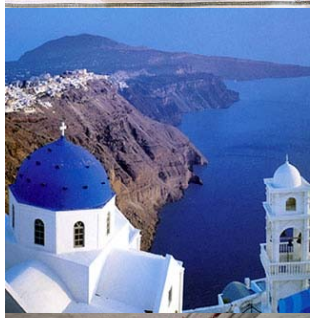


Specific Rates

- **Specific rates**
  - **refers to particular subgroups of the population defined by sex, age, race (ethnicity)**
    - **Eg. cause specific rates; age specific rates**

    See Tables 2.5 and 2.6

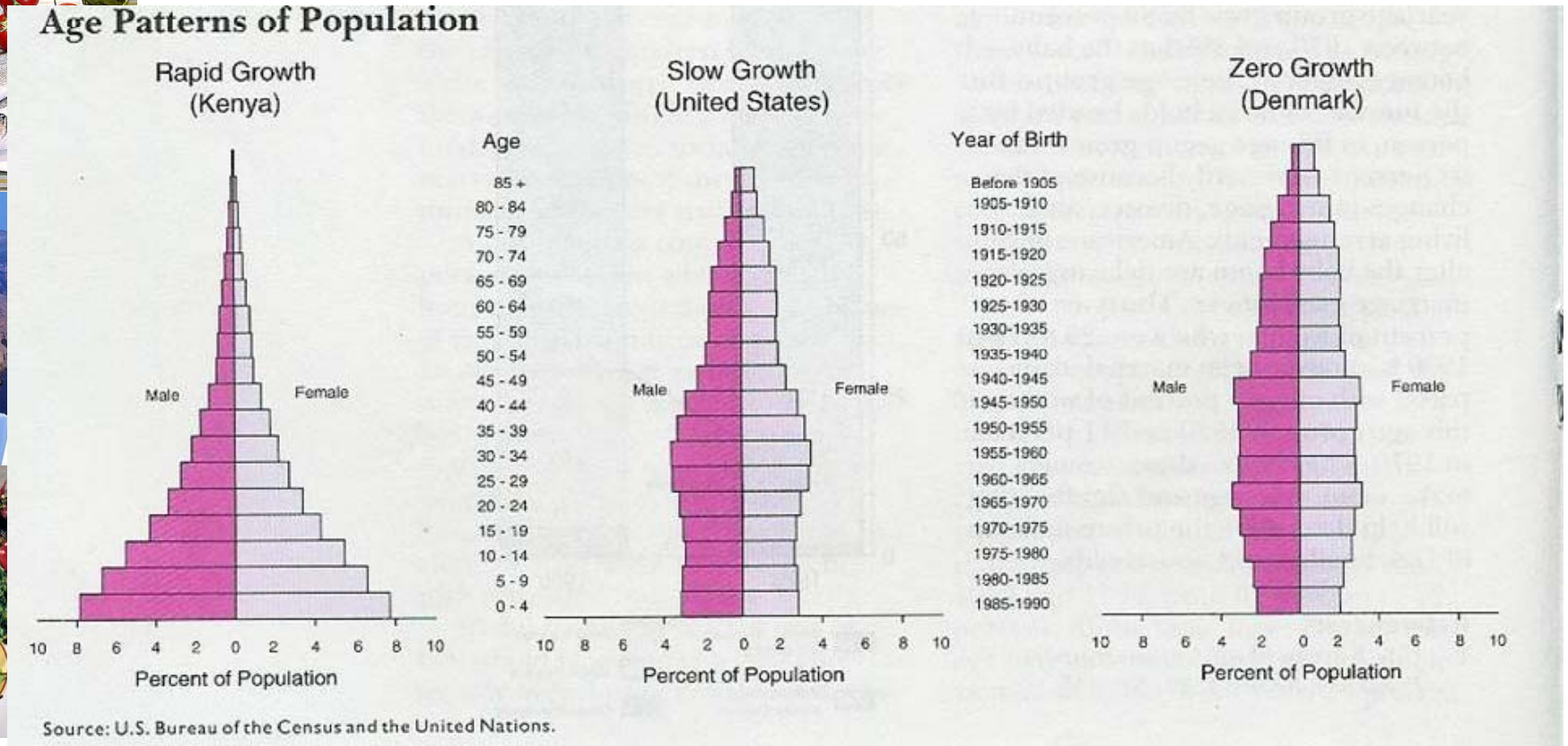
    Using a standard metric to adjust for pop. Composition factors – different age structures e.g. developed countries have more people in the older age groups

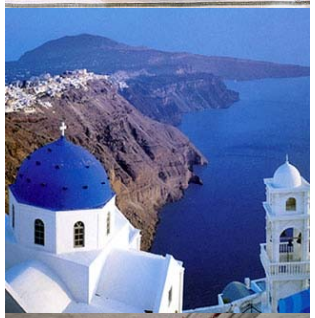


Standardised (adjusted) Rates

- **Standardised mortality/disease rates**
  - **summary measures of rate of morbidity or mortality in a pop^n where statistical procedures are applied to remove the effects of difference in composition of various pop^{n}s**
    - **Common adjusted factor - usually age**
  - **2 methods of adjustment (see Appendix 1, Webb et al. 2005)**
    - **<u>direct</u> - OK if age specific death rate known & the age distribution of the standard population.**
    - **<u>indirect</u> - used if age-specific death rate of population are unknown or unstable (eg. small populations)**
- **Standardised data ensures comparisons are not biased by changes such as age, race, sex**

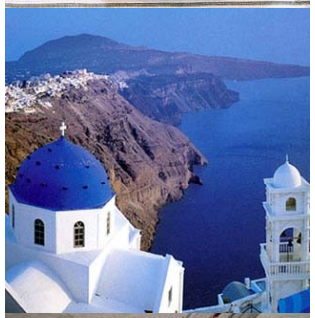
Shape of Population Affects Validity of Comparison of Mortality and Morbidity Rates Across Populations





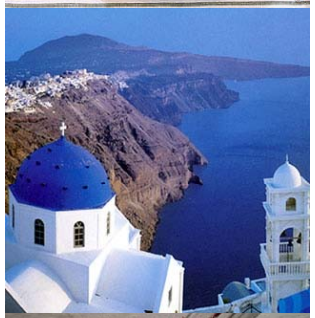
Is the reporting accurate?

- **There were 1941 cases of cancer in Northern Sydney during the first year of compulsory cancer registration in NSW (1972). In 1990, there were 2721 cases of cancer in the same region.**
- **Does this indicate that there has been an epidemic increase of cancer over this time period in Northern Sydney?**
- **What measures of incidence would you prefer to use to determine whether the rate of cancer was increasing?**



References

- Webb et al. 2005, Essential Epidemiology: An introduction for students and health professionals.
- UNSW Centre for Public Health 2005, Epidemiology for public health
- http://www.bmj.com/epidem/epid.2.html
- http://www.sjph.net.sd/files/vol4i4/SJPH-vol4i4-p414-417.pdf

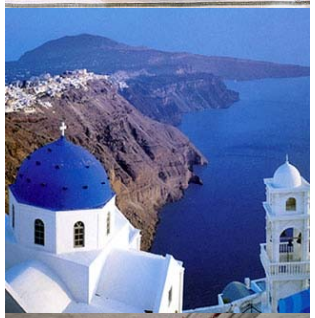


Part B: Assessment of Risk

A/Prof Jan Sansoni

6469: Epidemiology and Principles of Research

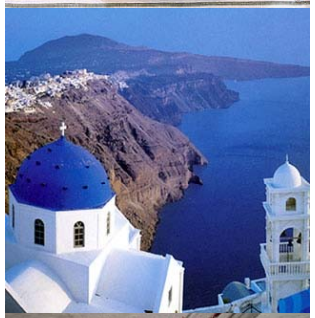
Acknowledgements: Gabrielle O'Kane, Vicki Deakin



Objectives for Part B: Assessment of Risk

At the end of this lecture students will be able to:

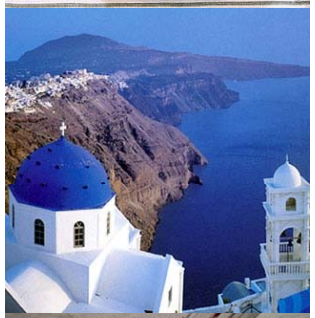
1. Explain the reasons for applying the measures of disease frequency in different population groups.
2. Draw up a 2 x 2 table of a disease and exposure to specific factors
3. Write an equation to describe relative risk (rate ratio) and odds ratios and calculate these with given data
4. Explain which studies use relative risk and odds ratios and why they are suitable in such studies
5. Write an equation to describe:
    1. attributable risk;
    2. attributable risk percent; and
    3. population attributable risk; and calculate these values with given data.



Why measure effect (association)?

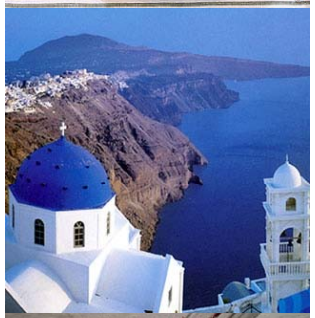
- Measuring disease in sub-populations defined by age, sex, geography or other factors might help detect patterns in disease.
- What might an epidemiologist be looking for?

Ref: UNSW Centre for Public Health, 2002



Measuring association

- First considerations:
  - Define those who have the risk factor and those who don't have the risk factor under investigation (the independent variable e.g. smoking)
  - Define those who have and do not have the disease (the dependent variable e.g. lung cancer)
  - These are commonly presented in the 2x2 table
    - Study/exposure/risk factor
    - Outcome factor = disease

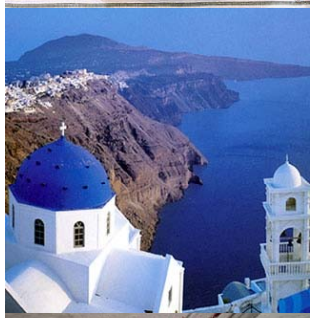


Assigning groups in epidemiological surveys *('2 x 2' table)*

DV=Disease (lung cancer)

IV= Exposure/ risk factor (smoking)

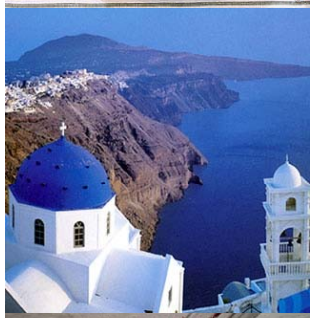
	Yes	No	Total
Yes	A	B	A+B
No	C	D	C+D
	A+C	B+D	N



Categories of association

- There are two broad categories of association
  - Ratio –RR tells us how much higher the rate is for 1 group vs. another (e.g. exposed group has 3* rate of disease)
  - Difference measures e.g. how many extra strokes among smokers –potential value of intervention

Ref: UNSW Centre for Public Health; Webb et al.2005 P90



Ratios - Relative differences in risk

- Rate Ratio or Relative Risk or Risk Ratio (RR)

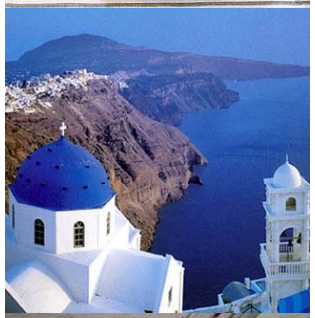
  $= \frac{\text{incidence (or prevalence) rate in exposed group}}{\text{incidence (or prevalence) rate in unexposed group}}$

  $= \frac{A/(A + B)}{C/(C + D)}$

  - The range of a relative risk is 0 to infinity
  - Relative risk of more than 1 implies that the exposure is positively associated with disease and may be a cause of disease (eg. High visceral fat and type 2 diabetes)
  - Relative risk of less than 1 implies that the exposure is negatively associated with disease and may be a preventative factor (eg. Physical activity and obesity)

  What does a relative risk of 1 mean?

Ref: UNSW Centre for Public Health



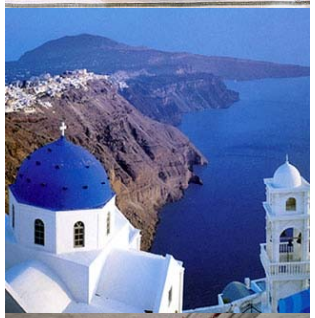
Ratios - Relative differences in risk

- An approximation to Relative Risk is the Odds Ratio (OR)
  - OR =Odds of exposure$_{(diseased)}$ /Odds of exposure (not-diseased)
  - Therefore, OR = [(A/C)/B/D)]

    = AD/BC

    Odds of case exposed = numerator

    Odds of control exposed = denominator

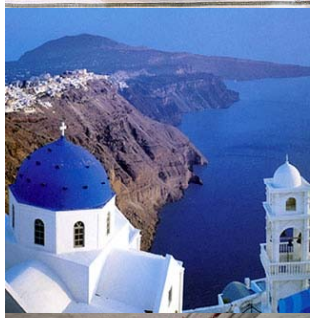


Odds Ratio used for Case Control Studies

Disease

Exposure/risk factor	Case	Control	Total
Yes	A	B	A+B
No	C	D	C+D
	A+C	B+D	N

Odds Ratio = AD / BC

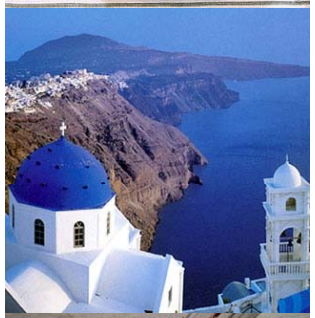


Practice example

Lymphoma

	Yes	No	Total
Gluten containing diet: Yes	5	100	
Gluten containing diet: No	2	150	

Calculate relative risk and odds ratio. How close are they?

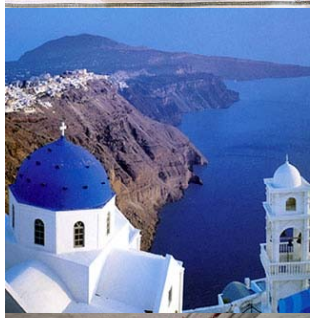


Relative Risk & Odds Ratio

- RR= $\frac{A/(A + B)}{C/(C + D)}$ $= \frac{5/105}{2/152} = \frac{.0476}{.0132} = 3.61$

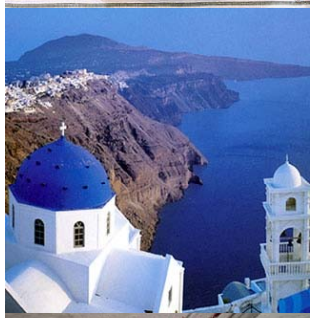
  OR = AD/BC $\frac{5x150}{100x2} = \frac{750}{200} = 3.75$

  These are quite similar when the disease is rare <5%. See the discussion on P109-P110. If the disease is more common then the relationship between RR and OR depends on how controls are selected for the study (Webb et al. 2005)



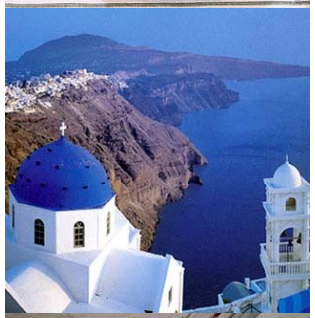
Associations in different study types?

- Cohort (follow up/longitudinal) studies
  - The frequency (#) of disease in the exposed subjects is compared to the # of disease in the unexposed
- Case-control studies
  - The frequency (#) of risk factor (exposure) in the cases (diseases) is compared to the # of risk factor (exposure) in the controls (non-diseased)
  - Hence, which ratio is used in cohort studies and which ratio is used in case-control studies?



Difference measures

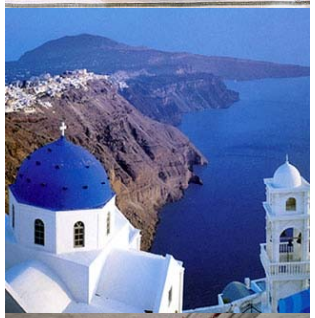
- *Risk difference* = Incidence (or prevalence) rate in exposed – Incidence (or prevalence) rate in unexposed
  - This is also known as *Attributable* risk (excess risk) for cohort studies
    - Equation
      - AR = I_E - I_U
        - » I_E = incidence rate of the disease in exposed group
        - » I_U = incidence rate of disease in unexposed group
      - Or AR = [A/(A+B)]- [C/(C+D)]
        - » Use previous example to work out attributable risk (lymphoma)



Difference measures continued…

- The range of a risk difference (or attributable risk) is –infinity to +infinity.
  - A positive risk difference implies that the disease is more common in the exposed than the unexposed.
  - A negative risk difference implies that the disease is less common in the exposed than the unexposed.
  - A risk of 0 implies the rate of disease is equal in the in exposed and unexposed – is equivalent to the relative risk (RR) of 1
- Risk Difference is better than relative risk for the assessment of the impact of an exposure as it quantifies the number of excess cases associated with that exposure

Ref: UNSW Centre for Public Health



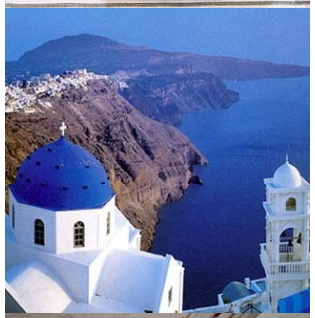
Difference measures continued…

- *Attributable risk percent* measures the proportion of risk in an *exposed group* that is due to the exposure. It is equivalent to the percentage of disease cases in exposed individuals that would be removed by eliminating exposure.
  - Equation
    - $AR\% = (I_{(exposed)} - I_{(unexposed)} / I_{(exposed)}) \times 100$

      or

      $= (RR-1/RR) \times 100$ (using relative risk)

The previous relative risk values for the lymphoma example are provided on the next slide

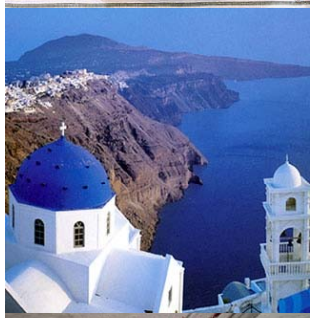


Practice example

		Lymphoma	
	Yes	No	Total
Gluten containing diet	Yes	5	100
	No	2	150

- Att.Risk = [A/(A+B)]- [C/(C+D)]
- = 5/(5+100) – 2/(2+150)
- =0.048 – 0.013
- =0.035

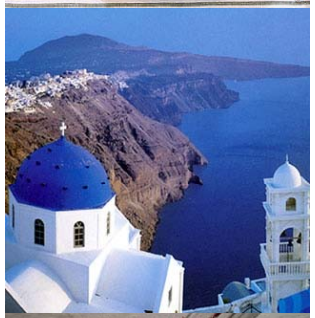
Att.Risk% = Ie-Iu/Ie
= 0.048-0.013/0.048 x 100
=0.72 x 100
=72%



Attributable Fraction

- In case control (vs. cohort studies) you use the Attributable Fraction based on the Odds Ratio instead
- $AF = \frac{OR-1}{OR} \times 100$

As usually you cannot calculate the attributable risk associated with disease for case control studies –use OR as the estimate instead (Slide 44)



Measures of health impact in the population

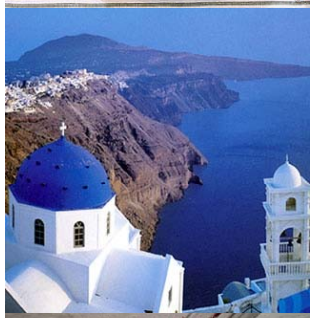
- Attributable risk and attributable risk percent tell us about the increased rate of disease in the exposed group (from a given study).
- May be useful to estimate the impact of the exposure on the population as a whole.
  - Population attributable risk (PAR) measures the change in incidence that removal of the exposure would have on the rate of disease in the entire population (not just the exposed).
  - <u>Equation:</u>

    PAR = Attributable risk x Prevalence of exposure

    $= (I_E - I_U)\ p$ or

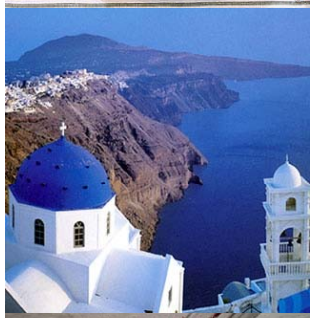
    = Incidence rate in population – Incidence rate in unexposed group

Ref: UNSW Centre for Public Health



Measures of health impact in the population

- Population attributable risk percent
  - This is known as aetiologic fraction.
  - Measures the proportion of total disease incidence that is attributable to the exposure and hence could be eliminated by removing the exposure.
  - Equation:
  - PAR% = $\frac{\text{Population attributable risk}}{\text{Incidence rate in population}} \times 100$



Ways of expressing the same risk of lung cancer associated with smoking

Measures of Smoking and Lung Cancer Deaths in Adult Males in the USA, 1986

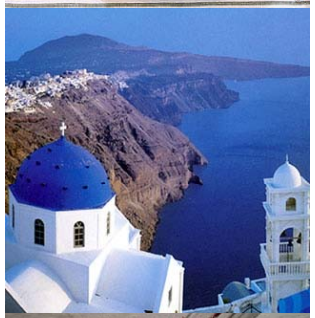
Measure	Amount
*Lung cancer deaths among smokers	191 per 100,000 per year
*Lung cancer deaths among nonsmokers	8.7 per 100,000 per year
†Proportion exposed (*Pe*) to the risk factor (effective population of smokers, averaged over time)	35%, or 0.35
†Population risk of lung cancer death	72.5 per 100,000 per year
†Relative risk, or RR	22 [191/8.7 = 22]
Attributable risk, or AR	182.3 per 100,000 per year [191 − 8.7 = 182.3]
Attributable risk percent in the exposed, or $AR\%_{(exposed)}$	95.4% [182.3/191 × 100 = 95.4]
Population attributable risk, or PAR	63.8 per 100,000 per year [72.5 − 8.7 = 63.8]
†Population attributable risk percent, or PAR%	88% [63.8/72.5 × 100 = 88]

* These rates were calculated from the data marked with a dagger.
† The source of data for measures shown with a dagger is as follows: Centers for Disease Control. Chronic disease reports: deaths from lung cancer—United States, 1986. Morbidity and Mortality Weekly Report 38:501–505, 1989.

Source: Katz, 1997

Your turn: allocating groups

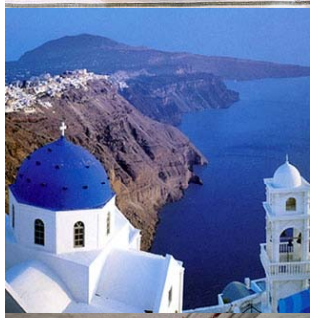
- You assemble 2 groups of 250 young men based on a past history of alcohol misuse and prospectively determine the rate of car accidents
  - You exclude subjects with a prior history of car accidents
- Over a period of 5 years, 135 of the exposed group and 38 of the unexposed group have car accidents, the rest have no car accidents
  - What type of study is this?
  - Can you assign these groups into a 2x2 contingency table?



Assigning groups into a *'2 x 2' table)*

Accidents

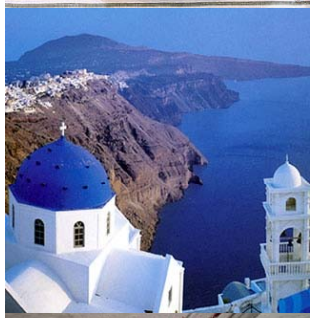
	Yes	No	Total
Yes	A	B	A+B
Alcohol intake No	C	D	C+D
	A+C	B+D	N



Assigning groups into a *'2 x 2' table)*

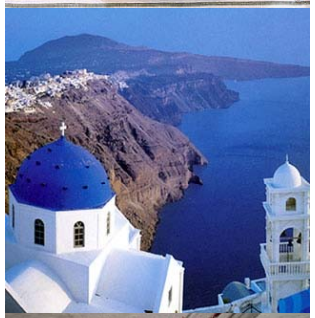
Accidents

Alcohol intake	Yes	No	Total
Yes	135	115	250
No	38	212	250
	173	327	500



Calculating risks

1. What is the relative risk of having a car accidents for the alcohol consumers?
2. Among alcohol consumers, what percentage of total risk for a car accident is due to the consumption of alcohol?
   - (ie calculate the attributable risk % in those who consume alcohol)



The relative risk of accidents in the alcohol consumers (answer)

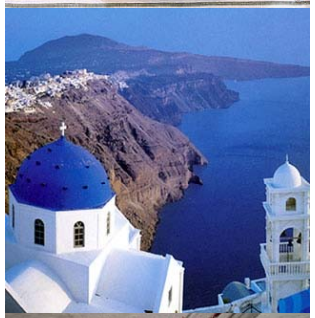
1. RR= [A/(A+B)]/[C/(C+D)]

$$= \frac{135/250}{38/250}$$

$$= 0.54/0.152$$

$$= 3.6$$

– How do you interpret this finding?



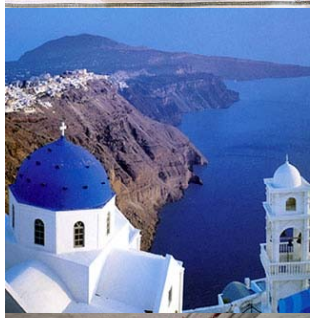
Attributable risk percent (answer)

2. Among alcohol consumers, what percentage of total risk for a car accident is due to the consumption of alcohol?
   - The % of total risk for an exposure (car accident) among those with the exposure (alcohol consumption) is the attributable risk percent or AR%
     - This can be calculated using the RR value

       AR % = RR-1/RR x100

     - AR% = 3.6-1/3.6 x100 = 72.2%

       In case control studies the Attributable Fraction (AR) based on the Odds Ratio is used instead.



References

- Webb et al. 2005, *Essential Epidemiology: An introduction for Students and Health Professionals*, Cambridge University Press, Melbourne, chapters 7 & 8.