



# Definition of Health Statistics

- Aggregated data describing & enumerating attributes, events, behaviours, services, resources, outcomes or cost related to health, disease and health services.
- Derived from surveys, medical records or administrative documents
- Vital statistics are a subset of health statistics
- Presented in various ways rates, ratios, proportions etc.

# Outline

- Rates
- Ratio
- Proportions (percentages)
- Morbidity-rates, incidence, prevalence
- Mortality-examples
- Crude and Adjusted rates
- Numerator/denominator
- Population pyramids

Numerator /
Denominator Issue

Numerator (the number above)
Denominator (the number below)

Rate, Ratio & Proportion - Diff				
Index	Numerator	Denominator		
Proportion	People with the disease	All people (with & without the disease)		
Ratio	People with the disease	People without the disease		
Rate	People with the disease in a given period	All people (with & without the disease) in a given		



# Rates, Ratios and Proportions (percentages)

- The need to assess the amount of growth, disease, disability, injury and death in population.
- Find ways to measure nominal data.
- The need to analyse, study factors affecting these measures of health status.
- are commonly used in public health planning reports, administrative decisions and planning of services.

Ra	tes				
Incidence Rate and Mortality Rate of Communicable Diseases Per 100,000 Population, 2006					
COMMUNICABLE DISEASES	INCIDENCE RATE	MORTALITY RATE			
AIDS HIV Infections (All forms)	6.91 21.88	3.66 1.43			



#### Rates

- Can compare mortality, morbidity and related health status occurrences to population groups of 100, 1,000, 10,000 or 100,000.
- To compare the data between one population group and another.
- Used in descriptive statistics and epidemiology.

9



#### Rates

- Is the amount or number of one thing measured in units of another.
- Similar to proportion except rates comes with a multiplier.
- Within a time period, usually in a year.
- The unit size is presented as 1000, 10 000, 100 000 (also called a base).
- E.g. Homicide rate (1993) in US was 8.7 per 100 000 population. The homicide rate in Malaysia (1.76 per 100,000), which is lower than that of Singapore (1.77 per 100,000)

10



#### Rates

- 3 basic factors needed:
  - □ Needs a numerator (the number of cases with the disease).
  - □ a denominator (the population of that area) and
  - □ a time period (usually a year).
  - □ Formula = <u>a</u> x Base

a + b

■ Rate = Number of cases X 1,000 (constant)
Population of the area
in a specific time period

..



#### Rates

#### 3 types;

- 1. Crude rates
- 2. Adjusted rates
- Specific rates morbidity & mortality, case fatality



#### Crude Rates

- Based on the number of events/experiences that happen in a total population in a certain period of time.
- Types usually used.
  - □ Crude death rates.
  - □ Crude birth rates total birth are used in the denominator.

3

## #

#### Limitations of crude rates

- Ignore information from subgroups/ special circumstances i.e. diff of demographics.
- Fail to show differences found.

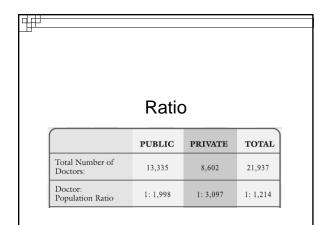
14

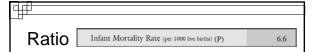


# Adjusted rates

- Use mathematical manipulations.
- Used to compare 2 or more groups, but differ in risk, or a 3<sup>rd</sup> variable is present that confuses presentation (confounder), data adjustment is needed.
- Use adjust for age, race, religion, sex, marital status etc.
- Direct /indirect method (discussed further towards the end of this lecture).

15



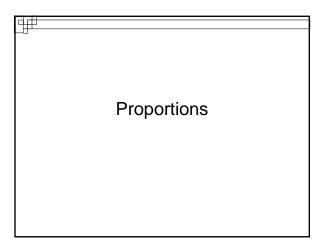


- Number of observations (a) in a group with a characteristic (such as having MI) divided by the number (b) of observations without the given characteristic (not having MI)
- Formula Ratio = a/b
- Sometimes rates are actually ratios i.e. Infant Mortality Rate (infants who die in the year of interest, but were born in the previous year, are counted in the numerator but not in the denominator).
- Relation in number, degree, or quantity existing between two things. Expressed as a fraction.

# Ratio

- Less useful than rates in descriptive statistics.
- Numerator is not included in the population defined by the denominator.
- E.g. 5 cases of measles compared to 30 children without measles. 5/30=0.166
- No time period needed.
- Not unlike percentage, ratio can exceed 100%.

Maternal Mortality Rate (per 1000 live births) (P) 0.3





# **Proportions**

- Is the number (a) of observations with a given characteristic (have MI) divided by the total number of observations (a+b) in a given group (such as taking
- Is a ratio in which the numerator is included as part of the denominator.
- Range 0.0 to 1.0.
- Difference between a ratio and a proportion is that the numerator of a proportion is included in the population defined by the denominator.
- No time period needed.

# **Proportions**

- Also called percentages (%) if measured in each hundred.
- Will give a clear view if presented properly. Can also be misleading.
- Formula: a/(a+b)

21

# Percentages

#### Environmental Health 2006

Percentage of Houses Served with Safe Water Suppy (Rural): 95.27 Percentage of Houses Served with Adequate Sanitary Latrines (Rural): 98.26

Percentage of Houses Served with Satisfactory Sullage Disposal (Rural): 59.52

Percentage of Houses Served with Satisfactory Waste Disposal (Rural): 68.00

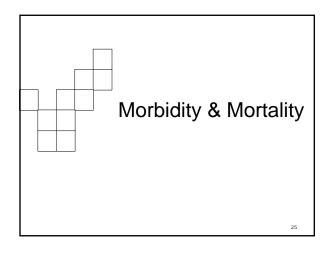
# Percentages

- Is proportion multiplied by 100%.
- Percentages should always be compared with the total numbers from which they come.
- - □ 80% participated from 1,200 elderly people.
  - 40 children are currently ill with measles out of 80 children in that pre-school.
    - 40 currently ill/80 in the pre-school = 0.5

    - In percentages = 0.5 x 100 =50%
      (0.5 is the proportion. 50% is the percentage)

# Percentages

- Should always state the total number i.e. N along with its percentage.
- E.g. 670 out of 750 persons or 89.3% of the participants was immunized.





- Means "death" or describes death and related issues.
- Mortality statistics are reported from information contained in death certificates, registries.
- Death rates can be presented by age, sex, cause of death.
- Affected by under-reporting and accuracy of diagnosis

26

Table 1. Statistics for Malaysia, 1998 (5)	_
Total population	21 446 031
Males	10 877 024
Females	10 569 007
Medically certified deaths	43 514
Uncertified deaths	54 392
Percentage of medically certified deaths	44.4%>
Common causes of medically certified deaths:	
Diseases of the circulatory system other than cerebrovascular disease	8848
Accidents, poisonings and violence	6564
Diseases of the respiratory system	4573
Malignant neoplasms	4498
Cerebrovascular disease	3367
Septicemia	2923
Others	12 741
Total	43 514

# Sources of Morbidity Statistics

- Hospital records/health services
- Notification of diseases
- National registries i.e. Cancer/HIV
- Special Morbidity Surveys i.e. NHMSI,II&III
- Data from blood banks/labs
- Absenteeism & medical leave

28

# Morbidity-Rates and Ratios

- Any disturbance in body function is considered a disease.
- Disease, illness, injury, disorders and sickness are all categorized under the single term-Morbidity.
- Morbidity is stated in a defined population.
- Usually expressed in terms of specific rates of incidence or prevalence.

Incidence &
Mortality Rate of
Communicable
Diseases per
1000 population
in 2006

	COMMUNICABLE DISEASES	INCIDENCE RATE	MORTALITY RATE
	AIDS	6.91	3,66
	HIV Infections (All forms)	21.88	1.43
	Chancroid	0.00	0.00
	Cholera	0.89	0.01
	Dengue Fever	64.37	0.01
	Dengue Haemorrhagic Fever	4.10	0.25
	Diphtheria	0.00	0.00
	Dysentery (All Types)	0.39	0.01
	Food Poisoning	26.04	0.01
	Gonococcal Infections	1.90	0.00
	Leprosy	0.89	0.00
	Malaria	19.87	0.08
	Measles	2.27	0.00
	Plague	0.00	0.00
	Poliomyelitis (Acute)	0.00	0.00
	Rabies	0.00	0.00
	Relapsing Fever	0.00	0.00
	Syphilis (All Forms)	3.06	0.01
	Tetanus Neonatorum	0.04	0.01
	Tetanus (Adult)	0.06	0.00
	Tuberculosis	62.56	5.37
	Tyhoid and Parathyphoid	0.77	0.02
	Typhus	0.06	0.00
	Viral Encephalitis	0.09	0.01
	Viral Hepatitis (All Forms)	9.37	0.18
	Whooping Cough	0.02	0.00
	Yellow Fever	0.00	0.00
	Hand, Foot and Mouth Disea	sc 19.30	0.02
- 1	Ebola	0.00	0.00



## a) Incidence

- The number of new cases with disease which comes into existence within a certain period of time per population at risk at beginning of the time interval.
- Should always be expressed of a unit of time.
- Incidence rate=number of new cases of a disease in
  a given time period
  Number of persons exposed to risk
  of developing the disease in the same time period

31



#### Denominator for Incidence

- Will change with time when numbers at risk move in or out.
- Thus, pick the population number/cases at a midpoint in the time period to represent the average population at risk.
- Population not at risk will be excluded from the denominator, such as;
  - □ individuals with the disease
  - who had the disease (if they are no longer susceptible)
  - □ or those who are not susceptible due to immunizations.

32



### Numerator for Incidence

- Role to provide specific information about the occurrences of a disease.
- Number of new cases starting within a time period.

33



## Concepts

- To determine the incidence of a disease, the population or group must be studied prospectively e.g. cohort study.
- Purpose- to ascertain the extent and rate of new cases occurring.
- Plotting of new cases based on time/date of occurrence will produce an epidemic curve.

34



#### Some Principles of Incidence Rate (IR)

- Can be used to estimate the probability of or risk of developing a disease during a specific time period.
- As incidence rate goes up, the probability of getting the disease goes up.
- Time-if the IR is higher during specified time of year (such as winter), the risk of developing the disease also goes up e.g. influenza.

Ħ

#### Some Principles of Incidence Rate (IR)

- Place-if the IR is constantly higher among people who live in a certain place, the risk of developing the disease goes up.
- Person-if IR are constantly higher among individuals with a specific lifestyle factor, the risk also goes up.
- Higher incidence implies many new cases, thus risk increases.
- When incidence of disease is known to be high, the existence of or potential of an epidemic becomes known and predictable.



# b) Prevalence

- Is the number of cases of disease, infected persons, or conditions, present at a particular time (not new cases).
- Prevalence of a chronic disease is high compared to its incidence.
- Point prevalence and period prevalence rate

37



### Prevalence

■ Period prevalence rate

number of existing cases of disease

= ----- within a time period x 1000 Average study population

■ Point prevalence rate

number of existing cases of disease

= ----- at a point in time x 1000

Total study population

38



# Affecting factors

- As a new disease develops in a population, incidence goes up. Prevalence goes up.
- The duration of a disease affects prevalence. When a disease has a long duration, prevalence remains higher longer.
- Intervention and treatment have effects on prevalence.

39



#### Prevalence rate

- Assess the total number of people in a group or population who have a disease at a specific time
- Prevalence rate=Incidence rate x average duration of disease.
- Controlled by 2 elements;
  - $\hfill\Box$  The number of individuals who have had the disease in the past.
  - ☐ The length and duration of the illness.

40



### Prevalence rate

Prevalence Number of existing cases

Rate = of the disease at a point in time x 1,000

Total population at a given time

#

# Comparative factors affecting prevalence rates

#### Rates increased;

- Immigration of ill cases.
- Emigration of healthy persons.
   Immigration of susceptible cases or those with potential of becoming cases.
- Increase in incidence.
- Prolongation of life of cases (increase of duration of disease).

#### Rates are decreased;

- Immigration of healthy persons.
- Emigration of ill cases.
- Improved cure rate of cases.
- Increased death rates from the diseases
- Decrease in occurrence of new cases.
- Shorter duration of disease.
- Death.

42

# Differences of Incidence & Prevalence

Incidence

Prevalence

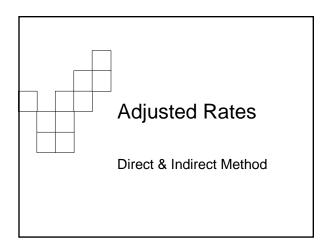
- New cases only.
- Prospective studies.
- More difficult to measure.
- Old and new cases.
- Cross sectional study.
- Easier to measure through a survey.

43

# Point prevalence

Is the number of cases (individuals) with a disease/condition/illness at a single point in time.

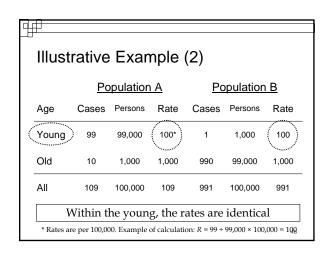
44

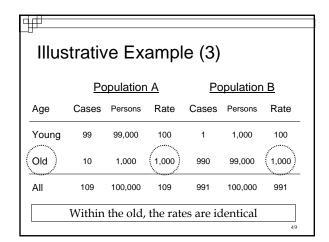


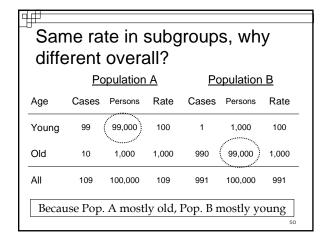
# Goals

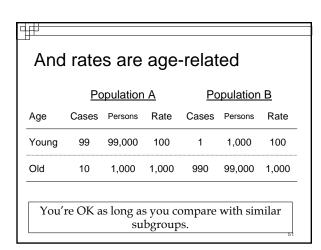
- To reduce distortions and incomparability of rates when making comparison over time and among populations
- To encourage "like-to-like" comparisons

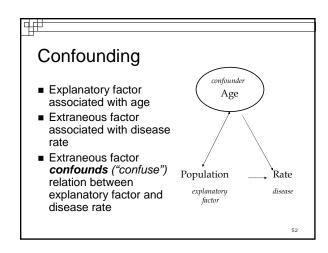
₩—						
Numerical Illustrations (1)						
Population A Population B						
Age	Cases	Persons	Rate per 100,000	Cases	Persons	Rate per 100,000
Young	99	99,000	100	1	1,000	100
Old	10	1,000	1,000	990	99,000	1,000
All	109	100,000	109	991	100,000	991
Overall rate of disease in Population B (991/100,000) is 9× that of Population A (109/100,000)						

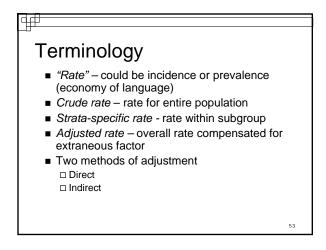


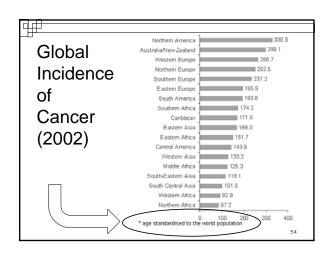














#### Methods

#### Direct

- Factor to measure, must categorize them.
- For each category (in each population), must know the specific rates.
- Use reference population as "standard".
- Able to make direct comparison.

#### Indirect

- Use when sometimes specific rates are not known.
- Results in standardized mortality ratio
   (defined as the number of observed deaths divided by the number of expected deaths)
- Result in the form of ratio.
- Make relative comparison.

55

# #

### **Direct Method**

- Recalculate the expected death rate using the demographic distribution of a standard population.
- Standard population
  - □ Population of one of the countries
  - □ Combination of countries
  - □ Artificial population

56



### General Idea, Direct Adjustment

- Apply strata-specific rates from study to a reference ("standard") age distribution
- Adjusted rate is a weighted average of strata-specific rates (with weights from reference population)

57

# U.S. **Standard** Million, 1991

#### Standard Million Age range 0 - 476,158 5 - 24286,501 24 - 44325,971 45 - 64185,402 65 - 7472,494 75+ 53,474 Total 1,000,000

58

### #

### Formula

$$aR_{direct} = \frac{\sum N_i r_i}{\sum N_i}$$

#### where

 $N_i$  = population size, reference population, strata i

 $r_i$  = rate, study population, strata I

Note: caps denote reference pop. values, while lower case denotes study pop. values

#### Florida & Alaska Mortality Example

■Crude rates (per 100,000)

 $\Box cR_{Florida} = 1026$ 

 $\Box cR_{Alaska} = 387$ 



# Age-Specific Rates (per 100,000)

i	Age	Alaska	Florida
1	0 – 4	214	238
2	5 – 24	80	64
3	24 – 44	172	208
4	45 – 64	640	809
5	65 – 74	2538	2221
6	75+	8314	6887

•Crude rates (per 100,000):  $cR_{Alaska} = 387$ ;  $cR_{Florida} = 1026$ 

61

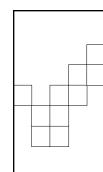
7						
Direct adjustment of Alaska rate						
i	Age	Rate	Std Million	No. of deaths		
		$r_i$	$N_i$	$N_i \times r_i$		
1	0 – 4	214	76,158	16,297,814		
2	5 – 24	80	286,501	22,920,080		
3	24 - 44	172	325,971	56,067,012		
4	45 – 64	640	185,402	118,657,280		
5	65 - 74	2538	72,494	183,989,772		
6	75+	8314	53,474	444,582,836		
		$\Sigma \rightarrow$	1,000,000	842,514,792		
	$aR_{direct}$ =	$= \frac{\sum N_i r_i}{\sum N_i}$	$=\frac{842,514,792}{1,000,000}$	≈ 843		



## Comparing Adjusted Death Rates

- Direct adjustment of Florida mortality rate using same standard million derives aR<sub>Florida</sub> = 784 (try to calculate it yourself using Excel to understand it better)
- cR<sub>Alaska</sub> = 387
- vs vs
- $cR_{Florida} = 1026$
- aR<sub>Alaska</sub> = 843/
- aR<sub>Florida</sub> = 784
- Conclude: slight advantage goes to Florida

63



indirect adjustment (NR for calculations)

64



## **Indirect Method**

- Can be done either by using the Age Specific Death Rate of the standard population or by using the age specific population of the standard population.
- Then get the ratio of Observed/Expected
- **480/532 = 0.90**
- Death Rate of males in Company < Standard Population

#

# Indirect Age-Adjustment

- Same goal as direct adjustment
- Based on multiplying crude rate by Standardized Mortality Ratio (SMR)

$$SMR = \frac{A}{\mu}$$

7nhord

A = observed number of cases in study population  $\mu$  = the expected number of cases (next slide)



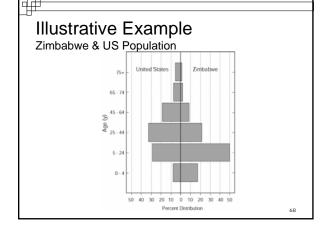
## Expected Number of Cases (μ)

$$\mu = \sum R_i n_i$$

where

 $R_i$  = rate, reference population, strata i  $n_i$  = population size, study population, strata i Recall: caps denote reference pop. values and lower case denote study pop. values

This is number of cases expected in study population if it had reference population's rates



# Indirect adjustment of Zimbabwe rate

i	Age	US Rate	Zimb Pop	Product
		$R_i$	n <sub>i</sub>	$R_i \times n_i$
1	0 – 4	.00229	1,899,204	4,349
2	5 – 24	.00062	5,537,992	3434
3	24 – 44	.00180	2,386,079	4,295
4	45 – 64	.00789	974,235	7,687
5	65 – 74	.02618	216,387	5,665
6	75+	.08046	136,109	10,951

 $\Sigma R_i n_i = 36,381$ 

$$\mu = \sum R_i n_i = 36,381$$
  $SMR = \frac{A}{\mu} = \frac{98,808}{36,381} = 2.72$ 

# 罪

#### Zimbabwe SMR

- Observed 98,808 deaths in Zimbabwe
- Expected 36,381 (based on US rate)
- SMR = 98,808 / 36,381 = 2.72
- Interpretation: Zimbabwe mortality rate is 2.72× that of US after adjusting for age

70

# F

# **Indirectly Adjusted Rate**

$$aR_{indirect} = (cR)(SMR)$$

- Zimbabwe crude rate = 886 (per 100,000)
- $\blacksquare aR_{\text{indirect}} = (886)(2.72) = 2340$
- c.f. to US rate of 860

#

# Adjustment for Multiple Factors

- We can adjust for multiple factors simultaneously through stratification
  - □e.g., 5 age groups, 2 genders, 3 ethnic groups
  - $\rightarrow$  5 x 2 x 3 age-gender-ethnic strata
  - $\hfill\Box$  Use direct or indirect method of adjustment



# Case Fatality Rate

■ Case fatality rate =

Number of deaths due to a disease

Number of people with the same disease

- Usually expressed in percentage.
- E.g. In Bergen county, there were 500-HIV positive people. 5 died within a year after diagnosis.
- CFR= 5/500 x 100 = 1%

# Annual crude mortality rate

■ All deaths during a year x 1000 Total midyear population

# Infant mortality rate

- Number of infant deaths less than one year of age x1000 Total number of live births during the same year
- Excludes stillbirths.

# **Neonatal Mortality Rate**

- Number of neonatal deaths x 1000 Total number of live births during the same year
- Neonatal period is interval between birth and 28 days postpartum.

# Perinatal mortality rate

- Number of perinatal deaths x 1000 Total number of live births during the same year
- Perinatal period is between 22 weeks of gestation to 4 weeks post partum.

# Maternal mortality rate

- Number of maternal deaths x 100,000 Total number of live births during the same year
- Pregnancy related deaths.



## Age-specific mortality rate

■ Number of people who died in a particular age group

x 1000

Total midyear population of the same age group during the same year



# **Fertility Rates**

- Crude Birth Rate
- General Fertility Rate
- Age Specific Fertility Rate
- Total Fertility Rate
- Gross Reproduction Rate
- Net Reproduction Rate
- Women-Child Ratio



## Crude Birth Rate (CBR)

#### =Total live births X 1000 Mid-year population

•Birth rates ranging from 10-20 births per 1000 are considered low, while rates from 40-50 births per 1000 are considered high.

•High birth rates can cause stress on the government welfare and family programs to support a youthful population, educating a growing number of children, creating jobs for these children when they enter the workforce, and dealing with the environmental effects that a large population can

·Low birth rates can put stress on the government to provide adequate senior welfare support system and also the stress on families to support the elders themselves. There will be less children or working age population to support the constantly growing aging population.



# General Fertility Rate (GFR)

X 1000 = Total live births women of childbearing age (15-44 or 15-49)

measurement of fertility rates only involves the reproductive rate of women, and does not adjust for the sex ratio

•For example, if a population has a total fertility rate of 4.0 but the sex ratio is 66/34 (twice as many men as women), this population is actually growing at a slower natural increase rate than would a population having a fertility rate of 3.0 and a sex ratio of 50/50. •This distortion is greatest in India and Myanmar, and is present in

### Age-Specific Fertility Rate (ASFR)

Total live births among =women age (x +n) X 1000 **Total women** age (x + n)

n =age range of 5 years

- An accurate rate.
- •Too many rates since we have 6 or 7 age groups
  •Different age proportions.

# Specific Fertility Rate by;

- ■Marital status
- ■"Birth order"



# Total Fertility Rate (TFR)

- $5 \Sigma_{x n} F_x X 1000$
- •The TFR is a measure of the fertility of an imaginary woman who passes through her reproductive life subject to all the age-specific fertility rates for ages 15-49 that were recorded for a given population in a given year.
- •The TFR represents the average number of children a woman would have were she to fast-forward through all her childbearing years in a single year, under all the age-specific fertility rates for that year.
- •In other words, this rate is the number of children a woman would have if she was subject to prevailing fertility rates at all ages from a single given year, and survives throughout all her childbearing years.



# **Total Fertility Rate**

- "Period TFR"
- •Current •Cross-sectional
- Hypothetical population •Easy to calculate
- •Commonly used
- "Cohort TFR"
- obtained by summing the age-specific fertility rates that actually applied to each cohort as they aged through time.
- Difficult to calculate
- •Real life experience

### Gross Reproduction Rate (GRR)

 $n \Sigma$  (ASFBR) X 1000

5 Σ number of daughters X 1000 born to a woman aged (x+n) Total women by age (x+n)

#### = TFR/ Gender ratio

• the average number of daughters that would be born to a •if 1000 women delivers more than 1000 daughters, there would be a generational increase.



# Net Reproductive Rate (NRR)

- Rate of Generation Replacement
- Generational Replacement Rate

X 1000 =  $n \Sigma_x (ASFBR)_x$ \*(women's survival)

If NRR=1, equal replacement If NRR > 1, population growth If NRR=1, stable population If NRR < 1, population decrease



# Net Reproductive Rate (NRR)

- NRR takes into account the survival of women in the reproductive age.
- (Women's survival is calculated using a life table, therefore need good death data)
- GRR > NRR
- If there is a large difference between the two, then the death rate of fertile women is high.

#### NRR Advantage

- Accurate
- •Measures generational replacement

- NRR Disadvantage
  •Need good data sources Only applicable in
- developed countries.

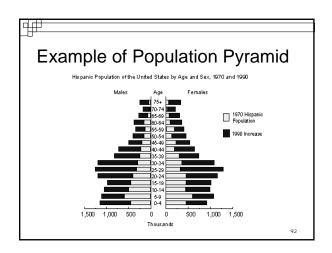
#### Child-Women Ratio

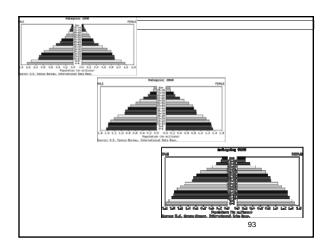
- \_ Total child below 5 years old Total women in the fertile age (15-44 or 15 to 49).
- The ratio of the number of children under 5 to the number of women 15-49
- Used by countries with poor birth registration
- Based on surveys
- Bias: Do not include children that died.

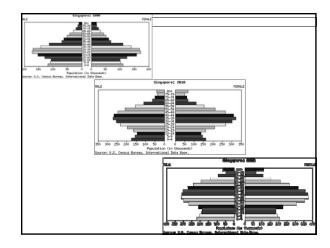


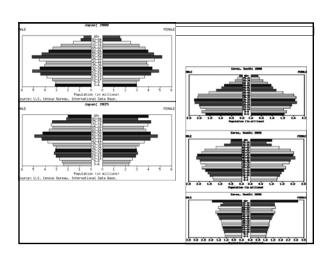
# Age and population pyramids

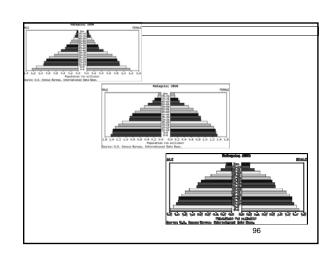
- To track changes in the population age groups over time.
- Changes may be affected by birth rates, fertility levels (number of 15-45 years old female), wars, death rates and migration.
- Male age groups usually left side, female in right side.

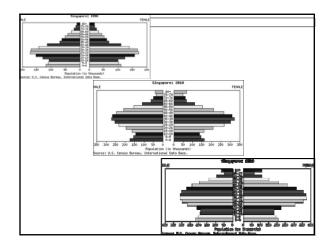


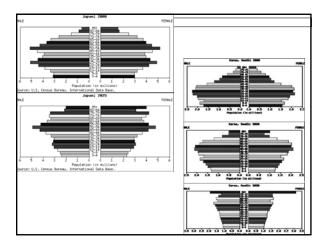


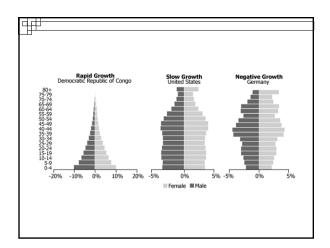


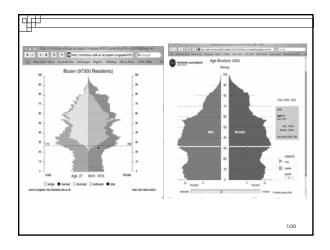












- Shape is caused by the numbers of persons born in the years of the age groups.
- Changes affect the gender and size of a population in different age groups.
- Pyramid uses 5 year incremental age, top is the oldest, bottom is the youngest.
- Usually 18 different age groups.

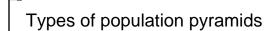
# Types of population pyramids

- Tall pointed

  Population with high birth rate, high death rate. Few people are in the older age cohort.
- Public health, socio economic and medical care conditions that do not allow cohorts to live into old age.
- Broad based-many babies born, IMR high.

Beehive shaped

Population with low birth rates, low death rates.

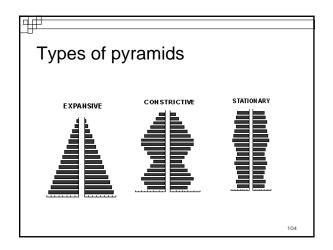


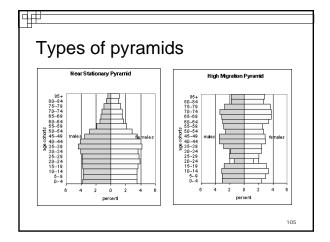
#### Hourglass

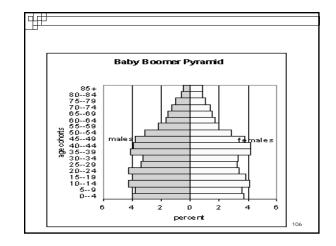
Transitional type population moving through the population due to low levels of births in the cohort of that time period. Blocked shaped

- A more industrialized society.
- Effective public health measures, good socioeconomic and medical conditions.
- High life expectancy.

103







# References

- An introduction to Epidemiology. 3<sup>rd</sup> edition. 2002. Thomas C.Timmreck. Bartlett Publishers.
- Kaedah Epidemiologi. 1990. Osman Ali.Dewan Bahasa dan Pustaka.
- Appleton and Lange's Review of Epidemiology & Biostatistics For The USMLE. 1994. Edward J. Hanrahan, Gangadhar Madupu. Appleton & Lange.
- Basic and Clinical Biostatistics. 2<sup>nd</sup> edition. 1994. Beth Dawson-Saunders, Robert G.Trapp. Prentice-Hall International Inc.