Data sources for chronic disease

Mortality Data from WHO Member States

- less than 100% completeness of deaths data;
- significant differences in coding practices across countries and in the use of 'ill-defined' codes;
- described at http://www3.who.int/whosis/mort/text/validated.cfm;

Mortality Data from WHO Member States

- For cardiovascular diseases in particular, even for countries with complete vital registration data, there are two main issues in estimating cause-specific mortality rates:
  - differences across countries in the proportion of CVD deaths that get coded to so-called garbage codes (mainly heart failure, and sudden death, also heart disease ill-defined codes).;
  - changes in reported rates associated with ICD revisions;

Data sources

A. Disease registers

Examples

⇒ Tuberculosis, cancer, renal failure, epilepsy, schizophrenia, diabetes, hypertension, thalassaemia, cystic fibrosis

Issues

⇒ Completeness

⇒ Data quality

⇒ Represents whole population?

Data sources

B. Population surveys

1. Surveys of <u>self-reported</u> impairment, disability, recent illness, risk factors

2. <u>Measurement</u> surveys
   - Blood (HIV, Hepatitis, diabetes)
   - Stool (helminths)
   - Other physiological tests (spirometry, ECG, vision)
   - Clinical examination (guinea worm, amputations)
   - IQ testing, neuro-developmental tests, motor function

Data sources

B. Population surveys — issues

1. Representativeness

⇒ Sampling: whole country?, specific age groups?, gender or ethnic bias?

⇒ Response rate

⇒ Specific groups excluded: homeless, illegal immigrants, institutionalised persons

Data sources

B. Population surveys — issues

2. Validity
   - ⇒ Does the survey measure what it set out to measure?
   - ⇒ "Tried" method
3. Reliability
   - ⇒ How stable are measurements over time? (Repeatability, reproducibility)

Data sources

B. Population surveys - Example 1

Household survey of mental disorders in Australia: prevalence of heroin addiction < 2 per 1,000.

However: Estimate based on contact with treatment facilities/justice : around 7 per 1,000

Reasons

1. Drug addicts often not in normal households

2. Self-reported drug use $\neq$ actual drug use

Data sources

B. Population surveys - Conclusions

Useful for important disabling conditions (mental disorders, COPD, musculoskeletal conditions, hearing loss, vision loss, cognitive function, diabetes, helminths, malnutrition, HIV)

But:

⇒ Query self-report

⇒ Examine selection bias

⇒ Give more weight to measurements even if only in selected group

Data sources

C. Epidemiological/cohort studies

⇒ Longitudinal studies: natural history of disease e.g. incidence, average duration, levels of severity, remission, case-fatality

⇒ Studies of chronic disabilities: mental retardation, acquired brain injury

⇒ Studies of particular population groups: ethnic minorities, homeless people, rural/urban differences

Issue

⇒ Generalisability of studies to own setting

Data sources

C. Epidemiological studies : review published and unpublished data

- Database searches (Medline, PsychLit, others)
- Select most relevant papers from abstracts
- Get and read most relevant papers
- Construct table of extracted info sorted by measure, e.g. prevalence, incidence, duration, severity etc.
- Contact authors for additional analyses or try to obtain unit record data

Data sources

Literature review - Asthma in Australia

Author	Place of study	Year of study	Age group	Wheeze last 12 months	Wheeze last 12 months + AHT
Bauman (1992)	Sydney	1990	5-12	22.4%	
Bauman (1992)	Eastern Australia	1990	5-12	19.5%	
	Melbourne			18.5%	
Peat (1994)	NSW 2 sites	1982	8-10		4.5% - 6.6%
		1992		23 - 27%	12.0% - 9.4%
Peat (1994)	NSW 2 sites	1991-92	8-10	22 - 26%	11.3% - 9.5%
Peat (1995)	NSW 7 sites	1991-93	8-11	24 - 38%	7.1% - 13.0%
Robertson (1992)	Rural Victoria	1990	7	23.6%	
	Melbourne	1990	7	23.1%	
			12	21.7%	
			15	18.6%	

Data sources

D. Health facility data

⇒ Limited use for burden of disease estimates

⇒ OK for conditions that generally lead to admission, e.g. in some countries: acute abdomen, MI, deliveries, more severe injuries

Issues

⇒ For most conditions not a proxy for health status

⇒ Episodes vs. persons

⇒ Accuracy of data sets: incentives and disincentives

Data sources

E. Sentinel Surveillance data

- ⇒ Limited use for burden of disease estimates
- ⇒ Good for trend monitoring, detection of epidemics, outbreak control

Issues

- ⇒ Under-reporting
- ⇒ Episodes vs. persons
- ⇒ Accuracy - incentives and disincentives
- ⇒ Definition of cases

Data consistency

Consistency - do the parameters make sense?

⇒ Between different parameters: e.g. incidence, prevalence, deaths, remission

⇒ Between different sources for the same bit of information: e.g. prevalence estimates

Data consistency

Adjusting for non-representativeness

- ⇒ Extrapolation of estimates for one particular region or ethnic group
- ⇒ Basis: epidemiological knowledge of determinants of health differentials (e.g. socio-economic status, rural/urban differences)

Data consistency

Obtain estimates even if you have little information?

Even if only expert "guestimates"

⇒ make an estimate!

⇒ Policy makers want to know

⇒ Uncertainty analyses of epi inputs

⇒ Often produces surge of studies/data

Data consistency

Expert consultation

⇒ Public health experts and clinicians

⇒ Discuss plausibility, other data sources

⇒ Get disease experts, opinion leaders interested to think about population health & resource allocation

⇒ But … expert opinion often biased: advocacy reasons, experts see “severe end of spectrum”

WHO tools for collecting comparable risk factor data

- STEPwise approach to NCD risk factor surveillance

- World Health Survey (WHS)

- Behavioural Risk Factor Surveillance System (BRFSS)

Additional data sources

- WHO databases
  - NATIONS tobacco database
  - WHO BMI/Obesity database
  - Global Alcohol database
  - WHO Mortality database and
  - CHOICE database on cost-effectiveness of medical interventions