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 self-reported impairment, disability, recent illness, risk factors

2. Measurement 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)
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 ≠ 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
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
C. Epidemiological studies: review published and unpublished data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

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