Research designs & bias in environmental epidemiology

1. challenges
2. designs
   • traditional
   • non-traditional
3. biases

Sverre Vedal

Environmental epidemiology: challenges

1. long latency
2. exposure measurement error
3. rare diseases
4. low-level exposure
5. small effect size
   notion of "more probable than not"

Morgenstern & Thomas, EHP 1993

calculation of RR in cohort study

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Death rate (deaths/person-yrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20 30 40 50 60 70 80 90</td>
</tr>
<tr>
<td>5 exposed subjects</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>5 unexposed subjects</td>
<td>1 1 1 1 1</td>
</tr>
</tbody>
</table>

exposed person-yrs = 20+30+40+50+60 = 200
unexposed person-yrs = 30+40+50+60+70 = 250

RR = \frac{5/200}{5/250} = 1.25
Environmental epidemiology: designs

A problem:
A public health worker in Turkey has been impressed by the large number of cases of mesothelioma in the region.
How can this impression be moved to a more intellectually (scientifically?) rigorous level?

---

types of study designs in environmental epidemiology

<table>
<thead>
<tr>
<th>traditional</th>
<th>non-traditional</th>
</tr>
</thead>
<tbody>
<tr>
<td>• case series</td>
<td>• time series</td>
</tr>
<tr>
<td>• &quot;ecological&quot;</td>
<td>• time series</td>
</tr>
<tr>
<td>• cross-sectional</td>
<td>• &amp; case-crossover</td>
</tr>
<tr>
<td>• case-control</td>
<td>• panel</td>
</tr>
<tr>
<td>• cohort</td>
<td>• quasi-experimental (ie, natural experiment)</td>
</tr>
</tbody>
</table>

• gene-environment interaction
• intervention?

---

Advantages of case series reports

• May identify new occupational or environmental hazard
  - *e.g., (besides asbestos and mesothelioma), vinyl chloride and angiosarcoma of the liver*

• Can lead to intervention
Limitations of case series reports

• Only anecdotal information
  "the plural of ‘anecdote’ is not ‘data’"

• May be a spurious cluster
  (Nassim Nicholas Taleb’s “Fooled by Randomness”)

“Ecological” study design

Approach
• correlate disease rates with geographical distribution
  of “exposure”

Example
• compare rates of mesothelioma of Turkish towns with
  environmental asbestos to those of other towns

Hasanoglu HC, et al. Int Arch Occup Environ Health 2005

Fig. 2 Scanning electron micrograph of tremellar asbestos taken
from the wall of a house (2000x magnification)

Hasanoglu HC, et al. Int Arch Occup Environ Health 2005

**"Ecological" study design**

<table>
<thead>
<tr>
<th>areas</th>
<th>mesothelioma cases (in 5 yrs)</th>
<th>population</th>
<th>incidence rate/100,000/yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arguvan, Hekimhan</td>
<td>7</td>
<td>52,004</td>
<td>2.7</td>
</tr>
<tr>
<td>rest of Malatya province</td>
<td>2</td>
<td>771,517</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Is there a problem?
“Ecological” study design

Main limitations

• no information on other risk factors for the disease (e.g., for lung cancer = age, smoking, occupational exposures)
• potential misclassification of people’s actual exposures

Cross-sectional design

Approach:
Describe/compare prevalence of some feature (e.g., disease/symptoms or physiological/imaging abnormality) typically by level of exposure
A “survey” or “slice in time” without regard to exposure or health outcome

Example:
Compare chest x-ray abnormalities in Turkish towns with environmental asbestos compared to other towns


Approach
Questionnaires and chest x-rays for all > 20 yrs old in village of Kureysler.

Findings
18% had pleural plaques and/or calcification c/w asbestos exposure.
Cross-sectional design


**Interpretation**

This is an example of a **descriptive** cross-sectional study in that it describes features of a population, but has no comparison group(s).

**More valuable** if study included comparisons* with other communities.

*To give the authors credit, they state that chest x-ray screening in other communities detected no pleural abnormalities, but no details are provided.*

Advantages of cross-sectional studies

- Good for less adverse outcomes (symptoms, physiological measures, imaging)
- Statistical power often good
- Direct contact with population sample (e.g., workers) permits additional individual-level data to be collected on:
  - other risk factors of the outcome
  - modifying factors [use of PPE, etc.]

Limitations of cross-sectional studies

- Very susceptible to selection bias (e.g., migration between jobs/regions influenced by health status).
  - Access only “survivor” population (e.g., workers most affected may have quit = "healthy worker effect").
- Whether exposure precedes the outcome (temporality) may be unclear.
types of study designs in environmental epidemiology

<table>
<thead>
<tr>
<th>traditional</th>
<th>non-traditional</th>
</tr>
</thead>
<tbody>
<tr>
<td>• case series</td>
<td>• time series</td>
</tr>
<tr>
<td>• &quot;ecological&quot;</td>
<td>&amp; case-crossover</td>
</tr>
<tr>
<td>• cross-sectional</td>
<td>• panel</td>
</tr>
<tr>
<td>• case-control</td>
<td>• quasi-experimental (ie, natural experiment)</td>
</tr>
<tr>
<td>• cohort</td>
<td>• gene-environment interaction</td>
</tr>
<tr>
<td></td>
<td>• intervention?</td>
</tr>
</tbody>
</table>

Time Series Studies

• Compare day-to-day changes in exposure with total event numbers in an area

• Group-level design: Total event counts are the outcome and a representative measure of exposure is the population exposure

• a short-term exposure effect design

removing long-term temporal trends

How is the Analysis Done?

- Aim: To relate daily event counts to daily exposure concentrations, adjusting for season and weather confounders
  - Disease model for daily event counts $Y_t$:
    \[
    \log(E(Y_t)) = f_1(\text{time}) + f_2(\text{weather}) + \alpha(\text{DOW}_t) + \beta(x_t)
    \]
  - $\beta$ is the parameter of interest for exposure $x_t$
  - $f_1, f_2$ are smooth functions
    - Nonparametric smooth $\Rightarrow$ GAM model
    - Parametric smooth $\Rightarrow$ GLM model
**NMMAPS: the essential (multi-city) air pollution time series study:**
short-term PM$_{10}$ mortality associations in 88 US cities

Dominici, Am J Epidemiol 2002

---

**Time Series Study Designs**

- **Strengths**
  - Data are typically easy to obtain
  - Individual-level factors are controlled by design
  - Generally good statistical power

- **Challenges**
  - Analytically complex
  - Exposure measurement error that differs for different exposures can complicate interpretation

---

**Case-Crossover Studies**

- **Concept:** The perfect control would be the case under a different exposure scenario (remember time machines)

- For transient exposures and outcomes with abrupt onset, each subject used as their own control

- Harvey Checkoway reviewed on March 27
Referent Time (control) Selection

- Critical in avoiding bias
- Might also want to match for temporal trends (day of week, month, season, etc.)
- Need to have rough understanding of timing so that select referent window outside of induction time or wash-out time

Panel studies

Longitudinal study where subjects (typically well-characterized) are followed (and measured) repeatedly over time

A variant of the repeated measures design except many more measurements than is typical of repeated measures

Example:

n=188 children ages 6-13 followed for 18 months with daily measurement of symptoms and peak flow


Panel studies: example

Vedal S et al.
Am J Respir Crit Care Med 1998; 157: 1034-1043
Panel studies

Advantages:
1. each subject their own control (so, no need to control for individual characteristics – potential confounders)
2. can get measured endpoints (and lots of them)

Disadvantages:
1. requires lots of resources to recruit and follow-up
2. analytically complex (e.g., time trends need to be controlled)

Quasi-experimental design (natural experiment)

Takes advantage of an unnatural intervention that is not randomized.

Examples:
1. pre-post study of water fluoridation comparing rate of dental caries in neighboring communities (Morgenstern & Thomas, 1993)
2. steel-mill strike reducing community air pollution levels (Utah Valley)

A steel mill strike in the Utah Valley: a natural experiment on PM and hospitalizations

Fig. 1. Monthly bronchitis and asthma admissions for all ages in Utah, Salt Lake, and Cache valleys, April 1980–1989.

Quasi-experimental design  
(natural experiment)

**Advantages:**
1. controls for many potential confounders  
2. often easy to carry out

**Disadvantages:**
1. since not really experimental, confounding is still a possibility

Environmental epi intervention design

**Example:**

- Randomly assigned Vit E and Vit C to 158 asthmatic children in Mexico City to assess differential response to air pollution
- Finding: lung function declined in association with increases in ozone and PM in the placebo group but not the treatment group

Case-control design

**Approach**
Compare past exposure of persons with disease (cases) with exposure of persons free of disease (controls).

**Example**
Compare history of residence in Turkish towns with high environmental asbestos in those with and without mesothelioma.
Design of case-control study

- relatively quick and inexpensive
- good for rare diseases or diseases with a long latency period
- can examine more than one exposure simultaneously

Advantages of case-control design

- inefficient if exposure is rare
- direct computation of incidence rates is not possible
- possible to confuse the temporal relationship between the exposure and the outcome
- particularly prone to bias (especially selection and recall "information" bias)
Cohort design

Approach
Compare rate of new disease in a well-defined population cohort according to degree of exposure.

Example
Compare mesothelioma rates in cohort of people from Turkish towns with varying degrees of environmental asbestos.

Design of cohort study

\[
\text{incidence rate (IR)} = \frac{\# \text{cases}}{\text{person-yrs}} \\
\text{IR (exposed)} = \frac{2}{25} = 0.080 \\
\text{IR (unexposed)} = \frac{1}{27} = 0.037 \\
\text{IR ratio (IRR)} = 2.16 = \text{RR}
\]
Advantages of cohort studies

- good when rare exposure
- when prospective, less chance of exposure information bias
- time relationship clear (especially if prospective)
- can examine more than one outcome
- directly measures incidence of outcome in exposed and unexposed (if internal comparison)

Limitations of cohort studies

- not good for rare diseases
- relatively costly and time-consuming (especially if prospective)
- if retrospective, often need records and the availability/quality of information may be a problem
- loss to follow-up may cause bias