An Overview of Analytic Epidemiology
A link between antecedent factors and some outcome—possibly a causal relationship, but not necessarily.

**Exposure (Risk Factor)**

- Exposures
  - “Risk factors”
  - Preventive measures
  - Management strategy
  - Independent variables

**Outcome**

- Outcomes
  - Dependent variable
  - Disease occurrence

**Examples:**

- Lack of exercise → Heart disease?
- Flu Shot → Dystonia Disorder?
Evolution of Information

Descriptive Studies
- Description; Hypothesis Generation
  - Case Report
  - Case-Series
  - Cross-Sectional
  - Correlational

Analytical Studies
- Hypothesis testing
  - Compare groups
  - Case-Control
  - Cohort Study
  - Clinical Trial (Intervention Study)

Observational Comparison Studies
- Evaluation of Intervention
In analytic studies one enrolls subjects from a population and groups them in some way to make comparisons that test association between risk factors and outcomes.
Two Basic Strategies for Testing Associations

• Cohort type of study

• Case-Control study
Cohort Type Studies

Compare Incidence

Case-Control Studies

Compare Prior Exposures
Surveillance system for reportable infectious diseases identifies a case of Salmonella food poisoning.

Subsequent surveillance and active case finding revealed a substantial number of recent cases.
Based on the descriptive epidemiology, it is clear that the parent-teacher luncheon is the source of the outbreak (presumably one of the food dishes). But which food dish was responsible?
The attendees of the luncheon constitute a well-defined group (cohort) that is the “source population.” Any of a number of food dishes could have been the “exposure” responsible for causing Salmonella in some members of the cohort.

An intuitive approach would be to ask all attendees in the cohort what they ate (their exposures). Then, for each food dish sort the attendees into those who ate it and those who did not, and then compare the incidence of Salmonellosis (the outcome) in the two exposure groups.
45 attendees completed the questionnaire which asked whether they had become ill and which dishes they had eaten.

- Among the respondents, 23 reported having eaten a cheese appetizer. 16 of these people became ill.
- 22 denied eating the cheese. 9 of these people became ill.

Was the cheese the culprit? Is there evidence of an association between eating the cheese appetizer (exposure) and developing Salmonellosis (outcome)?
### The Data – A “Line Listing”

<table>
<thead>
<tr>
<th>ID</th>
<th>gender</th>
<th>Salmonella</th>
<th>Cheese</th>
<th>Mushrooms</th>
<th>Pasta</th>
<th>Pot Salad</th>
<th>Veg Lasagna</th>
<th>Chick &amp; Rice</th>
<th>manicotti</th>
<th>Veggies</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>M</td>
<td>No</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>Yes</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>Yes</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>M</td>
<td>No</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>No</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>No</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>M</td>
<td>No</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>M</td>
<td>Yes</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>M</td>
<td>Yes</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>F</td>
<td>Yes</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>M</td>
<td>No</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>M</td>
<td>No</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>F</td>
<td>No</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>M</td>
<td>Yes</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>F</td>
<td>Yes</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>M</td>
<td>No</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>F</td>
<td>No</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>F</td>
<td>Yes</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>F</td>
<td>Yes</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>F</td>
<td>Yes</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>25</td>
<td>M</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>26</td>
<td>F</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>27</td>
<td>M</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>28</td>
<td>M</td>
<td>Yes</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>29</td>
<td>F</td>
<td>Yes</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>F</td>
<td>Yes</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>31</td>
<td>M</td>
<td>No</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Did those who were exposed to a given dish have a higher probability of disease compared to … … those who were not exposed?
Method #1 for sampling: identify exposed people & non-exposed people and compare their risk of disease. (Esp. useful for rare *exposures*, like asbestos.)

<table>
<thead>
<tr>
<th>Exposed</th>
<th>Sick</th>
<th>Not Sick</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>![Exposed Sick]</td>
<td>![Exposed Not Sick]</td>
<td>14</td>
</tr>
<tr>
<td>No</td>
<td>![Exposed Sick]</td>
<td>![Exposed Not Sick]</td>
<td>28</td>
</tr>
</tbody>
</table>

Risk in exposed = $\frac{6}{14} = 43\%$; risk in unexposed = $\frac{4}{28} = 14\%$
<table>
<thead>
<tr>
<th>Ate Cheese App.? (Exposed)</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>16</td>
<td>7</td>
</tr>
<tr>
<td>No</td>
<td>9</td>
<td>13</td>
</tr>
</tbody>
</table>

Total: Yes = 23, No = 22

Probability of illness (risk):
- Yes: $\frac{16}{23} = 0.70$
- No: $\frac{9}{22} = 0.41$
How Did the Risks Compare?

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ate Cheese App.? (Exposed)</td>
<td>Yes</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>9</td>
</tr>
<tr>
<td>Probability of illness (risk)?</td>
<td>16/23 = 0.70</td>
<td>9/22 = 0.41</td>
</tr>
<tr>
<td>Risk Ratio</td>
<td>0.70 / 0.41 = 1.71</td>
<td></td>
</tr>
</tbody>
</table>

Those who ate the cheese appetizer had 1.71 times the risk of developing Salmonellosis compared to those who did not eat the cheese appetizer.
They looked at each of the “risk factors” (exposures) with a separate 2x2 table. The summary of the results looked like this.
Did people who ate manicotti have a greater incidence of Salmonella compared to those who did not eat it?

Key question:
Did people with a particular “exposure” have a greater incidence (risk) of disease?
How many times greater was risk in those exposed to manicotti?

Risk Ratio = \frac{93\%}{6\%} = 16.7

“The risk of Salmonellosis was 16.7 times greater in people who ate manicotti compared to those who didn’t.”
The Cohort: People Attending the Luncheon

- - - - -

Manicotti

No Manicotti

Time passes

Compare Incidence

Sick
The source population was small and discrete (attendees of appreciation luncheon) & there was the ability to contact all members of the cohort or a substantial proportion of them.

They could list all foods served at the luncheon & ask each respondent which foods they ate & whether they got sick.

They could, therefore, determine the exposure status & outcome status for the majority of the cohort. So, they could calculate incidence and RR for each food item.

The disease was common; 58% of the cohort got it.
Between February 25 and 27, 2004 six cases of HAV infection in Marshfield residents were reported to MDPH. In addition, a case of hepatitis A in a Plymouth resident, employed in Marshfield, was reported.” (eventually there were 20 cases).

Marshfield had 1 case in 2002 and 0 cases in 2003.

“The increase in the number of reported cases … during February in a confined geographic area was an indication of a possible outbreak of hepatitis A infection.”
Abrupt onset: fever, malaise, anorexia, nausea, and abdominal discomfort; sometimes diarrhea. Jaundice may follow. May be asymptomatic. Infected humans (symptomatic or not) shed the virus into stools.

Transmission: fecal-oral route (ingesting the virus)
- food contaminated by an infected food worker
- produce irrigated/processed with contaminated water
- shellfish from contaminated water
- drinking feces-contaminated water
- sexual: (e.g., oral-anal contact).

Incubation period: 15–50 days (avg.= 28–30).

Most infectious from 1–2 weeks before symptoms until 1 week after.
Descriptive Phase

(generate hypotheses about the source)

- **Person**: characteristics?
- **Place**: specific locations or setting?
- **Time**: does it vary over time?
Based on these clues:

- Knowledge of biology of hepatitis A (transmission, incubation)
- Time course: epidemic curve of “point source”
- Diverse age, occupation, location
- Interview with a series of cases & similarities in restaurant use

They hypothesized that the source was probably an infected food handler at:

- Rick’s Deli
- McDonald’s
- Jaime’s Pub
- Papa Gino’s
- Friendly’s
Was it feasible to test these hypotheses with a cohort study?
**Hepatitis Outbreak** – Problems

- No clear cohort and only a small # of cases scattered across South Shore. *(rare outcome)*

- No obvious event/place that tied them all together. The source population was large & diffuse with unknown borders, and **only 20 cases** had been identified.

- They couldn’t interview all residents of MA South Shore.
Of the thousands of people exposed at the responsible restaurant, only a small % became ill. So if we took a random sample of people who ate at each restaurant, the incidence might be 0 even in the offending restaurant.
1. The disease is rare.
2. There are many exposed individuals, but most of these are not diseased.
3. Yet, the proportion of exposed individuals among the disease cases may be higher than the proportion of exposure among the controls. (There may be an association.)
If I somehow had exposure and outcome information on all of the subjects in the source population and looked at the association using a cohort design, it might look like this:

<table>
<thead>
<tr>
<th></th>
<th>Diseased</th>
<th>Non-diseased</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>7</td>
<td>1,000</td>
<td>1,007</td>
</tr>
<tr>
<td>Non-exposed</td>
<td>6</td>
<td>5,634</td>
<td>5,640</td>
</tr>
</tbody>
</table>

If we are calculating the risk ratio, the key information is the exposure distribution in the disease cases relative to the exposure distribution in the total population.
If we are calculating the risk ratio, the key information is the exposure distribution in the cases relative to the exposure distribution in the total population. And the exposure distribution in non-diseased people is similar to that in the total population.

<table>
<thead>
<tr>
<th></th>
<th>Diseased</th>
<th>Non-diseased</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>7</td>
<td>1,000</td>
<td>1,007</td>
</tr>
<tr>
<td>Non-exposed</td>
<td>6</td>
<td>5,634</td>
<td>5,640</td>
</tr>
</tbody>
</table>

\[
\frac{7/1007}{6/5640} = 6.53
\]

\[
\frac{7/6}{1007/5640} = 6.53
\]

\[
\frac{1.16667}{0.1785} = 6.53
\]
If the key information is the exposure distribution in the cases relative to the exposure distribution in the total population, then we could just take a sample of the non-diseased people in order to estimate the exposure distribution in the total population.

<table>
<thead>
<tr>
<th></th>
<th>Diseased</th>
<th>Non-diseased</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>7</td>
<td>10</td>
<td>?</td>
</tr>
<tr>
<td>Non-exposed</td>
<td>6</td>
<td>56</td>
<td>?</td>
</tr>
</tbody>
</table>

\[
\frac{7/1007}{6/5640} = 6.53
\]

\[
\frac{7/6}{10/56} = 6.53
\]

\[
\frac{1.16667}{0.1785} = 6.53
\]
In other words, if I want to estimate a risk ratio for a rare disease, it is more efficient to find cases, but then just take a sample of non-diseased “controls” in order to estimate the exposure distribution in the entire population.

\[
\frac{7}{1007} \quad \text{Risk Ratio} \quad \frac{7/6}{10/56} \quad \text{Odds Ratio}
\]

\[
\frac{7/1007}{6/5640} = 6.53 = \text{Risk Ratio} \quad \frac{7/6}{10/56} = 6.53 = \text{Odds Ratio}
\]
Method #2 for Sampling

Enroll diseased people & non-diseased people and compare their odds of having been exposed. (Esp. useful for rare outcomes, e.g., birth defects.)

<table>
<thead>
<tr>
<th>Exposure Status</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Sick</td>
</tr>
<tr>
<td>No</td>
<td>Not Sick</td>
</tr>
</tbody>
</table>

Odds of exposure = 6/4; odds of exposure = 8/24
With no defined cohort and a rare outcome, the case-control strategy is much more efficient:

Find as many _sick people_ (cases) as you can and ask them about all their exposures (where they ate). Then find _non-affected people_ (controls) and ask them about the same exposures.

You can’t measure incidence, but you can measure the odds of exposure to each restaurant in the cases (sick people) and compare to the odds of exposure in well people (controls).
Design:
• Find cases with disease; find non-disease ‘controls’.
• Compare the groups with respect to past exposures.

**Case-Control Study**

People with Hepatitis A (cases)

People without it (controls)

Assess Prior Exposures

Compare odds of eating at ....
Evaluating Multiple Possible Risk Factors

Odds of Eating at:

- Rick’s Deli
- McDonald’s
- Jaime’s
- Friendly’s

<table>
<thead>
<tr>
<th></th>
<th>Case Yes</th>
<th>Case No</th>
<th>Control Yes</th>
<th>Control No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>No</td>
<td>5</td>
<td>24</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>7</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td>32</td>
</tr>
</tbody>
</table>
Marshfield Hepatitis Outbreak

Ate at Rick’s Deli

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>No</td>
<td>1</td>
<td>29</td>
</tr>
</tbody>
</table>

18 ate at Rick’s 1 didn’t
Odds = 18/1

7 ate at Rick’s 29 didn’t
Odds = 7/29
Those who ate at Rick’s Deli had 75 times the risk of getting hepatitis A compared to those who did not eat there.

Odds of exposure: 18/1  7/29

Odds Ratio = \( \frac{18/1}{7/29} = 75 \)
## Results:

<table>
<thead>
<tr>
<th>Restaurant</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rick’s Deli</td>
<td>74.6</td>
</tr>
<tr>
<td>McDonald’s</td>
<td>3.5</td>
</tr>
<tr>
<td>Jaime’s Pub</td>
<td>2.4</td>
</tr>
<tr>
<td>Papa Gino’s</td>
<td>1.1</td>
</tr>
<tr>
<td>Friendly’s</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Cohort Type Studies

Exposed

Compare Incidence

Non-Exposed

Case-Control Studies

Exposed

Non-Exposed

Diseased

Non-Diseased

Compare Prior Exposures

Compare Odds of Exposure to Risk Factor
Heart disease is a chronic disease.

Which study design should we use?
A difference in incidence suggests that the exposure is associated with the disease.

We could use a cohort type of design.
Or we could use a case-control design.

Compare odds of exposure (inactivity).

“Were you inactive?”

People with CAD (cases)

People without CAD (controls)
Choice of study design will depend on:

- degree of existing knowledge,
- whether the outcome is rare,
- whether the exposure is unusual,
- resources, time, money.
The study is *planned & designed to answer questions* in a specific area. Non-diseased subjects meeting eligibility criteria are enrolled. Detailed baseline information on lifestyle & exposures is collected from each & they are followed over time.
Employees of a tire manufacturing company.

Do chemicals used in tire manufacturing increase risk of death?

This study was **not preplanned**. The investigator has to go back to pre-existing data that was not necessarily acquired in a precise, predetermined way. Follow up may have been incomplete.
Retrospective vs. Prospective Cohort Studies

Start of Study

Retrospective Cohort
- Risk factor +
- Risk factor -

Prospective Cohort
- Risk factor +
- Risk factor -

Compare disease incidence.

Past

Future
We need to understand determinants of cancer and CHD in women.

Enroll & assess exposures at the beginning.

After time has elapsed investigators use the prospectively collected data to answer many questions.

A Prospective Cohort Study

The Cohort

117,000 Nurses without cancer or CVD

Compare incidence of heart attack

Follow-up

Start of Study

Future
A Randomized Clinical Trial (Intervention Study)

Similar to a prospective cohort study, but the investigator assigns exposure (treatment).

Example: Does low-dose aspirin reduce risk of myocardial infarction (heart attack)?

Randomly assign subjects to a treatment or “risk” group.

Compare rates of MI, stroke, etc.
A Clinical Trial

22,000 male MDs without CVD

Does low-dose aspirin prevent heart attacks?

Start of Study

Enroll & assign exposure (treatment) at the beginning.

Aspirin

Placebo

Follow-up

After time has elapsed investigators use the prospectively collected data to answer many questions.

Compare incidence of heart attack
## Fatal Myocardial Infarction

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
<th>Incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspirin</td>
<td>10</td>
<td>11,027</td>
<td>11,037</td>
</tr>
<tr>
<td>Placebo</td>
<td>26</td>
<td>11,008</td>
<td>11,034</td>
</tr>
</tbody>
</table>

Risk Ratio (RR) = 9/24 = 0.38
**Case-Control**

- **Past**
  - Compare risk factor frequency.
  - Cases → controls

- **Future**

**Retrospective Cohort**

- **Past**
  - Compare disease incidence.
  - Risk factor +
  - Risk factor -

**Prospective Cohort**

- **Future**
  - Compare disease incidence.
  - Risk factor +
  - Risk factor -

**Clinical Trial**

- **Past**
  - Compare disease incidence.
  - Treated
  - Not Treated

**Start of Study**
**Cohort**

Disease-free subjects are enrolled and then grouped by their exposure; then compare incidence.

**Case-Control**

Find diseased subjects and a non-diseased comparison group; compare odds of exposure.
When reading a paper, it isn’t always clear what the study design is. Sometimes there is a combination of strategies. However, you should think about what the predominant design features are.

- Provides a framework for thinking about the study.
- Alerts you to weaknesses in some study designs.
Identifying the Study Design

Is it based on information about individuals?

Or averages in populations?

Correlational (Ecologic)
Identifying the Study Design

Is there just one group?

Did all subjects have the disease? (Case Series)

Did they evaluate presence of disease and risk factors at the same point in time? (Cross-sectional Survey)

Do you have heart disease?

Are you active?
Identifying the Study Design

Two or more groups being compared?

• How were they selected? Did they find people with disease [cases] and then find a comparison group without disease [controls]? (**Case-Control**)

  Compare past exposures

• Identify non-diseased people & group them by risk factor status? Then follow them longitudinally to compare incidence? (**Cohort Study**)

  Compare incidence over time
In prospective cohort studies conception, design, & enrollment occur *before* anyone develops the outcome.

**Prospective**
- Enroll non-diseased subjects; collect baseline exposure data
- Follow up at *intervals* to get accurate outcome data.
- Compare incidence over time

**Retrospective**
- Identify a cohort retrospectively (e.g. tire manufacturing workers vs. desk employees. Look at what subsequently happened to them.
- Compare incidence over time
Identifying the Study Design

Did the investigators assign subjects to a treatment or intervention and follow them to compare outcomes? (Clinical Trial)

Aspirin

Placebo

Compare incidence over time
Oral Contraceptives & Liver Cancer. Previous case reports of liver cancers in women on OCs. The authors contacted all cancer registries & collected information on all females with liver tumors.

<table>
<thead>
<tr>
<th>OC Use</th>
<th>Age Category 16 - 25</th>
<th>Age Category 26 - 35</th>
<th>Age Category 36 - 45</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>% 31</td>
<td>% 43</td>
<td>% 22</td>
</tr>
<tr>
<td>No</td>
<td>% 20</td>
<td>% 10</td>
<td>% 29</td>
</tr>
<tr>
<td>Unknown</td>
<td>% 49</td>
<td>% 48</td>
<td>% 49</td>
</tr>
</tbody>
</table>

1. Case series
2. Case-control study
3. Retrospective cohort
4. Prospective cohort
5. Randomized clinical trial

What kind of study was this?
<table>
<thead>
<tr>
<th>State</th>
<th>Annual per capita Tobacco Sales</th>
<th>Lung Cancer Mortality Rate in 1965/100,000 pop.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>$600</td>
<td>92</td>
</tr>
<tr>
<td>Florida</td>
<td>$450</td>
<td>75</td>
</tr>
<tr>
<td>Georgia</td>
<td>$500</td>
<td>80</td>
</tr>
<tr>
<td>North Carolina</td>
<td>$550</td>
<td>66</td>
</tr>
<tr>
<td>Virginia</td>
<td>$400</td>
<td>45</td>
</tr>
<tr>
<td>Alaska</td>
<td>$200</td>
<td>35</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>$150</td>
<td>33</td>
</tr>
<tr>
<td>New York</td>
<td>$175</td>
<td>20</td>
</tr>
<tr>
<td>New Jersey</td>
<td>$200</td>
<td>23</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>$250</td>
<td>22</td>
</tr>
</tbody>
</table>

1. Case series  
2. Case-control  
3. Retrospective cohort  
4. Cross-sectional survey  
5. Correlational (ecologic)
Dr. Villerme notes that mortality varies among districts in Paris. He tried to correlate mortality with the distance of the arrondissement from the Seine River, the relationship of the streets to the prevailing winds, the arrondissement's source of water and local climatological factors such as soil type, exposure to the sun, elevation and inclination of the arrondissement.
Villerme found that mortality correlated closely with the degree of poverty in the arrondissement (estimated as the % of people exempted from tax). The findings did not spark action.
Villerme found that mortality correlated with the degree of poverty in the arrondissement (estimated as the % of people exempted from tax). The findings did not spark action.

Type of study?

1. Case series
2. Case-control
3. Retrospective cohort
4. Cross-sectional survey
5. Correlational (ecologic)
<table>
<thead>
<tr>
<th>City</th>
<th>Median Household Income</th>
<th>Premature Deaths /100,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lynn</td>
<td>38000</td>
<td>470</td>
</tr>
<tr>
<td>Lowell</td>
<td>40000</td>
<td>466</td>
</tr>
<tr>
<td>Springfield</td>
<td>30000</td>
<td>459</td>
</tr>
<tr>
<td>Newton</td>
<td>89000</td>
<td>218</td>
</tr>
<tr>
<td>Brookline</td>
<td>68000</td>
<td>233</td>
</tr>
<tr>
<td>Barnstable</td>
<td>47000</td>
<td>275</td>
</tr>
</tbody>
</table>

![Graph showing a negative correlation between median household income and premature deaths per 100,000 population.](image.png)
What kind of study was this?

Investigators in Bergen, Norway sent questionnaires about respiratory health, allergies, smoking habits, and occupational respiratory exposures to a random sample of residents between the ages of 15-70. After two reminders, 2,819 responses were obtained. Of these, 1,646 reported exposure to tobacco smoke from other members of their immediate family.

1. Case series
2. Case-control
3. Retrospective cohort
4. Cross-sectional survey
5. Correlational (ecologic)
A study in N. Engl. J. Med. examined whether eating a Mediterranean diet had any association with mortality in Greek adults. A baseline questionnaire was used to determine how closely the subjects followed a traditional Mediterranean diet, and the group was followed for 2 years, during which they determined the cause of death among all subjects who died.

<table>
<thead>
<tr>
<th>Mediterranean Diet Score</th>
<th>Deaths in 2 yrs</th>
<th>Alive</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>High (close adherence)</td>
<td>44</td>
<td>2586</td>
<td>2,630</td>
</tr>
<tr>
<td>Medium</td>
<td>61</td>
<td>3747</td>
<td>3,808</td>
</tr>
<tr>
<td>Low (poor adherence)</td>
<td>74</td>
<td>2383</td>
<td>2,457</td>
</tr>
</tbody>
</table>

1. Case-control
2. Retrospective cohort
3. Prospective cohort
4. Randomized clinical trial

What kind of study was this?

A 43-year-old American woman developed a fever after traveling in Peru for 3 weeks. She visited Lima and Nazca and then traveled to the Sacred Valley of Urubamba, followed by Cuzco and Machu Picchu, where she hiked. She received numerous insect bites. Sixteen days after returning to the US she developed fever, insomnia, muscle aches, nausea, headache, and mild cough. At the hospital she was found to have anemia and an enlarged spleen (splenomegaly). Laboratory tests determined that her blood was infected with a genus of bacterium called Bartonella.

1. Case report
2. Case series
3. Case-control
4. Retrospective cohort
5. Clinical trial
6. Ecologic

What kind of study?
In 2003 a mass immunization against cholera was conducted in Beira, Mozambique. The following year there was an outbreak of El Tor Ogawa cholera in Beira. To assess the usefulness of the vaccine investigators compared the frequency of vaccination between persons with culture-confirmed cholera severe enough to have prompted them to seek treatment and age- and sex-matched neighborhood controls who did not have diarrhea.

1. Case series
2. Cross-sectional
3. Case-control study
4. Retrospective cohort
5. Prospective cohort
6. Clinical trial

Study type?

People who take analgesic drugs frequently may be at increased risk of chronic kidney failure. These authors used a kidney dialysis registry to find 716 patients with kidney failure; they randomly selected 361 subjects without kidney disease from the same geographic area. They used phone interviews to estimate their cumulative past use of analgesics and compared the two groups.

1. Case series
2. Case-control
3. Retrospective cohort
4. Prospective cohort
5. Clinical trial

What kind of study?

In 1976 the Nurse’s Health Study enrolled 121,700 female RNs who completed a mailed questionnaire regarding their medical history & lifestyle. The women have returned follow up information every two years. This study grouped them by exercise level & BMI and compared mortality rates among different levels of these two risk factors.

1. Cases series
2. Case-control
3. Retrospective cohort
4. Prospective cohort
5. Clinical trial

**Type of study?**
Glucosamine and chondroitin sulfate are orally administered substances that have been used for years to treat joint problems in horses. Since they are relatively non-toxic there has been increasing interest in them for treating osteoarthritis, but there is controversy about their efficacy. These investigators randomly assigned 1583 patients with symptomatic knee osteoarthritis to receive 1500 mg of glucosamine daily, 1200 mg of chondroitin sulfate daily, both glucosamine and chondroitin sulfate, 200 mg of celecoxib daily, or placebo for 24 weeks. The primary outcome measure was a 20 percent decrease in knee pain from baseline to week 24. The primary outcome measure was whether the patient achieved a 20 percent decrease in pain as measured by the WOMAC pain subscale, a standardized, previously validity tool for assessing joint pain.

What kind of study is this?