IDENTIFYING CAUSES OF DISEASE

PROBLEM BASED LEARNING EXERCISE
SNOW ON CHOLERA

Prepared by Prof. Ahmed Alkafajei

This learning exercise tells the story of John Snow and his brilliant investigations on cholera epidemics in the nineteenth century in London, England, to illustrate the use of epidemiology to identify causes of disease to help prevent the disease or control its occurrence.

Who is Snow?

Last's dictionary of Epidemiology (1983) introduces Snow as: [SNOW, JOHN (1813-1858) London general practitioner and early anesthetist, (he assisted Queen Victoria's delivery of two of her children with chloroform). His fame rests upon his observation, brilliant deductions, painstaking personal inquiries, and analytic studies of cholera outbreaks in the mid-19th century in London and elsewhere. All are recorded in "On the mode of communication of cholera" (London: Churchill, 2nd ed. 1855), which can be regarded as the first definitive working text on epidemiology and which also contained an explicit statement of the germ theory of transmission, written 30 years before Koch discovered the cholera vibrio].

Applying Present Day Terminology to Snow's Investigations

1. Descriptive study - Ecological study.
2. Hypothesis generation.
3. Hypothesis testing:
   * Case - control study.
   * Historical cohort study.
4. Public health application.

Snow's story does not follow the sequence adopted by the present learning exercise, nor does it use the present day epidemiological terms.

The aim of the learning exercise, being to illustrate the use of epidemiology in the identification of disease causation, dictated the sequence and mode of presenting Snow's findings in this exercise.

Snow on Cholera

The story of Dr. John Snow's brilliant investigations on cholera epidemics in London during 1849 - 1854, would be retold here utilizing the hypothetico - deductive approach.
The Descriptive Ecological Study on Epidemics of 1849 & 1853-54

Snow observed the following association based on group characteristics:

“Higher death rate from cholera in subdistricts that use Southwark & Vauxhall water as compared to those supplied by Lambeth company.”

Descriptive study - Ecological study - An example.

“The drainage from the cesspools found its way into the well attached to some houses at Locksbrook, near Bath, and the cholera making its appearance there in the autumn of 1849, became very fatal. The people complained of the water to the gentleman belonging to the property, who lived at Weston, in Bath, and he sent a surveyor, who reported that nothing was the matter.

The tenants still complaining, the owner went himself, and on looking at the water and smelling it, he said that he could perceive nothing the matter with it. He was asked if he would taste it, and he drank a glass of it. This occurred on a Wednesday; he went home, was taken ill with cholera, and died on the Saturday following, there being no cholera in his own neighborhood at the time.”

Ecological Evidence from 1853 Epidemics

Area supplied by both companies
Death rates per 100,000:

<table>
<thead>
<tr>
<th></th>
<th>Population</th>
<th>Deaths from Cholera</th>
<th>Cholera death rate per 1000 population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lambeth</td>
<td>182/314781</td>
<td>57.82</td>
<td></td>
</tr>
<tr>
<td>Southwark &amp; Vauxhall</td>
<td>374/467803</td>
<td>79.95</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>374/482435</td>
<td>77.52</td>
<td></td>
</tr>
</tbody>
</table>

There is an association as evidenced by the higher death rate for subdistricts that use Southwark and Vauxhall water as compared to Lambeth subdistricts.

Ecological Evidence from 1854 Epidemics

Districts with water supplied by

<table>
<thead>
<tr>
<th></th>
<th>Population</th>
<th>Deaths from Cholera</th>
<th>Cholera death rate per 1000 population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Southwark and Vauxhall Company Only</td>
<td>167,654</td>
<td>844</td>
<td>5.0</td>
</tr>
<tr>
<td>Lambeth Company Only</td>
<td>19,133</td>
<td>18</td>
<td>0.9</td>
</tr>
<tr>
<td>Both Companies</td>
<td>300,149</td>
<td>652</td>
<td>2.2</td>
</tr>
</tbody>
</table>

The previous 2 slides identify the experience of groups of people living in (districts) rather than the experience of individuals. This is why these findings are termed “ecological”.

The difference in mortality rates between districts supplied by Southwark-Vauxhall and those by Lambeth became clearer in 1854 because in 1854 mortality rates were computed for districts supplied by each company alone. In 1853 these rates were for districts supplied by both companies together. (Please see the map).
Try to answer the followings:

✍ What is your hypothesis about cholera transmission? Its rationale.

✍ What interpretation do you make of the ecological data? What additional information could you have sought at that time to further investigate the cause of the epidemic?

Snow’s Hypothesis

Cholera risk is related to the drinking of water supplied by a particular commercial company, and by inference to the source of water from which the company obtained its water.

Hypothesis Testing

Based on the epidemiologist uses a two stage sequence of reasoning:

(a) The determination of association between a characteristic (polluted water) and a disease (cholera).

(b) The derivation of biological inference from such a pattern of association.

The method to determine the statistical association falls into one of two broadly defined categories:

(a) Association based on group characteristics (ecological)

(b) Association based on individual characteristics.

Association between disease and characteristic

(a) Do persons with the disease have the characteristic more frequently than those without the disease? This usually is done by a case-control method of study.

(b) Do persons with the characteristic develop the disease more frequently than those who do not have the characteristic? This is done by a Cohort method of study.

Snow did both.

Hypothesis Testing

Case-Control Investigations

Broad street epidemic (Aug. - Sept.1854)

“Nearly all deaths had taken place within a short distance of the (water) pump of Broad street compared to almost cholera-free neighboring areas.”

And 5 more instances; All point to:

Persons with disease (cholera) have the characteristic (polluted water) more frequently than those without the disease.

Broad Street epidemic

“The most terrible outbreak of cholera which ever occurred in this kingdom is probably that which took place in Broad Street, Golden Square, and the adjoining streets. Within two hundred and fifty yards from the pump in that street there were upwards of five hundred fatal attacks of cholera in ten days.”
**Broad Street epidemic**

The following figure and table illustrate how John Snow made use of information on the time sequence and geographical location of cases.

A black dot for each death is placed at the location of the house in which a fatality occurred.

---

**Broad Street Outbreak, London, 1854.**

<table>
<thead>
<tr>
<th>Date</th>
<th>No. of fatal cases</th>
<th>Deaths</th>
<th>Date</th>
<th>No. of fatal cases</th>
<th>Deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aug. 29</td>
<td>1</td>
<td>1</td>
<td>Sept. 8</td>
<td>12</td>
<td>30</td>
</tr>
<tr>
<td>30</td>
<td>8</td>
<td>2</td>
<td>9</td>
<td>11</td>
<td>24</td>
</tr>
<tr>
<td>31</td>
<td>56</td>
<td>3</td>
<td>10</td>
<td>5</td>
<td>18</td>
</tr>
<tr>
<td>Sept. 1</td>
<td>143</td>
<td>70</td>
<td>11</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>116</td>
<td>127</td>
<td>12</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>54</td>
<td>76</td>
<td>13</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>46</td>
<td>71</td>
<td>14</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>36</td>
<td>45</td>
<td>15</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>37</td>
<td>16</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>28</td>
<td>32</td>
<td>17</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

---

**The widow at Hampstead district**

A widow had not been in the neighborhood of Broad Street for many months. A cart went from Broad Street to West End every day and it was the custom to take out a large bottle of the water from the pump in Broad Street, as she preferred it. The water was taken on Thursday 31st August, and she drank of it in the evening, and also on Friday. She was seized with cholera on the evening of the latter day, and died on Saturday.

A niece, who was on a visit to this lady, also drank of the water; she returned to her residence, in a high and healthy part of Islington, was attacked with cholera on the evening of the latter day, and died also.

There was no cholera at the time, either at Hampstead or in the neighborhood where the niece died.

---

**Readers tasks**

- What was the date of onset of the epidemic in Broad Street?
- What type of epidemic curve is suggested by the distribution of cases?
- What are the possible reasons for the termination of the epidemics?
- Do you think that the hypothesis of linking polluted water with cholera epidemic was strengthened by the findings from the Board Street outbreak? To what extent do you think that the cause of the outbreak was finally established?
The hypothesis incriminating polluted water in cholera transmission seems to be stronger now after it was substantiated by investigating Broad Street epidemic and its “Case-Control” Comparisons:

The investigation in Broad Street outbreak and the cited instances indicated that the water from the Broad Street pump was used by a much higher proportion of cholera victims (cases) than by those who escaped the disease (controls).

The proportion of deaths to 10,000 houses during the first seven weeks of the 1854 epidemic, July 8th-August 26th, South London.

<table>
<thead>
<tr>
<th>Water Supply of Individual Houses</th>
<th>Number of Houses</th>
<th>Deaths from Cholera</th>
<th>Deaths in each 10,000 houses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Southwark Vauxhall Company</td>
<td>40,046</td>
<td>1,263</td>
<td>315</td>
</tr>
<tr>
<td>Lambeth Company</td>
<td>26,107</td>
<td>98</td>
<td>37</td>
</tr>
<tr>
<td>Rest of London</td>
<td>256,423</td>
<td>1522</td>
<td>59</td>
</tr>
</tbody>
</table>

Mortality From Cholera in London Related to the Water Supply of Individual Houses in Districts Served by Both Companies July 8 to August 26, 1854

<table>
<thead>
<tr>
<th>Water supply</th>
<th>Population 1851</th>
<th>Deaths from cholera</th>
<th>Death rate per 1000 population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Southwark &amp; Vauxhall company</td>
<td>98,862</td>
<td>419</td>
<td>4.2</td>
</tr>
<tr>
<td>Lambeth company</td>
<td>154,615</td>
<td>80</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Public Health Application

Even though the actual disease agent was not known at that time (Snow had no knowledge of the existence of cholera vibrio) the epidemic was controllable, by controlling access to the contaminated water.

Hypothesis Testing
Quasi-Experiment with Cohort Analysis

During the epidemic of 1854 Snow conducted an experiment on the grandest scale, involving 300,000 people.

The supply of water to each individual house, as well as the number of fatal attacks from cholera in each house were ascertained.

Reader’s Tasks:

Try to answer the following questions:

What questions might be raised concerning the “Proof” of Snow’s finding?

Interpret the findings in tables 4.4 and 4.5. What is the difference in information conveyed by the two tables?

How could you quantify the strength of association between cholera deaths and source of water supply?
Questions that are raised concerning the “Proof of Snow’s finding.”

• Which variables were eliminated from consideration due to the “Natural experiment”?
• Was there random possibility that people were using a particular water supply?
• Complete enumeration of cases of the disease and population at risk?
• How accurate was Snow’s method for grouping people by their water source?
• How much possibility was there for misdiagnosis of cases or misclassification of the source of exposure for the population at risk?
• Could unapparent infections have affected the distribution?

Interpretations of the findings in 2 tables of cohort analysis

The event of the epidemic subdivided each group (exposed and unexposed) into two subgroups, those who developed cholera and those who did not. Because the participants of both exposed and unexposed groups were free from cholera at the beginning of the observation period (the time of the epidemic), it is clear that the suspected cause preceded the onset of the epidemic.

This study was very useful in “Proving” that a caused association existed. The “Relative Risk” (RR) can be calculated to measure the strength of association between the risk and the disease.

Interpretations of the findings in 2 tables of cohort analysis

The calculation of incidence rates is peculiar to cohort studies, and Snow’s investigation on 1854 epidemic is one such study. He ascertained the mortality experience among the exposed and unexposed people to the polluted water of Southwark-Vauxhall Company. However, both the exposure to the risk and the outcome (disease) happened in the past in regard to Snow’s time of investigation.

This type of cohort study is called retrospective (historical) cohort. Snow’s study is a cohort study even though the cohort (exposed and unexposed groups) assembled or formed in the past (but before the epidemic) according to their choice of the water supply (natural experiment).

Interpretations of the findings in 2 tables of cohort analysis

The difference in the results shown in 2 tables: The first table speaks about the incidence of mortality per 10,000 houses whereas second table expresses the mortality incidence per 1000 individuals.

The former table relates the cohort experience per group of people rather than individuals. It shows the results of cholera deaths per a unit of houses, it is an ecological finding, with all the disadvantages of ecological analysis. While second table expresses the cohort findings per individuals.

The results in second table are shown per unit of population, that lend itself readily for the calculation to measure the strength of association between the risk (polluted water) and cholera deaths, by estimating the RR.

Quantifying the risk in Snow’s experiment of July-August 1854 epidemic:

Relative Risk (RR) is defined as (example of Snow’s natural experiment):

Death rate from cholera among the exposed ÷ Death rate from cholera among unexposed

Attributable Risk (AR) is defined as:

Death rate from cholera among the exposed - Death rate from cholera among unexposed

The attributable risk percent is the attributable risk divided by the rate among exposed persons, expressed in percent:

Applying those measurements to Snow’s data second table) (of cholera epidemic of 1854

RR = Cholera death rate among those drinking water of Southwark-Vauxhall company (exposed) DIVIDED BY cholera death rate among those drinking water of Lambeth Company (Un exposed).

= 4.2 per 1000 population / 0.5 per 1000 population

= 4.238/ 0.517

= 8.2
Applying those measurements to Snow’s data of cholera epidemic of 1854

AR = cholera death rate among exposed MINUS cholera death rate among unexposed.
= 4.2 per 1000 population MINUS 0.5 per 1000 population.
= 3.7 per 1000 population.

Attributable Risk Percent = 3.7 per 1000 population / 4.2 per 1000 population x 100.
= 88.1%

These three indicators measures the risk from different perspectives

The RR shows that cholera deaths among the exposed people, to the polluted water of southwork - vauxhall co., is eight times that of the unexposed people who drinks water from Lambeth co. (8 - fold difference )

AR tells that 3.7 deaths per 1000 population are solely due to the polluted water of the southwark-vauxhall co.

Attributable risk percent portrays that 88% of cholera deaths could be prevented if the polluted water of the southwark-Vauxhall co. is purified or it’s source be changed to avoid contamination.

What about absolute risk?

The rates whose denominators are population at risk in general, are measures of ABSOLUTE RISK, that don’t distinguish illness among those exposed from illness among those unexposed to polluted water.

Thus, measures of absolute risk (Incidence of cholera in the population regardless of their water supply) do not provide a direct answer to one of epidemiology’s basic questions, which is, how much excess disease a factor such as polluted water, might produce in the population, or how much of the disease might be prevented.

Reader’s task

Are you satisfied that there is a causal relationship between polluted water and cholera transmission from result of Snow’s epidemiological study?

Epidemiologist’s Response

The answer to this question is divided into the following headings:
- Establishing association.
- Establishing practical significance.
- Establishing causality. The greatest challenge in epidemiology.
- Application of causal criteria to Snow’s study.

Establishing association:

Statistical methods can not establish proof of a causal relation when an association (Statistical) has been demonstrated. The causal significance of an association is a matter of judgment which goes beyond any statement of statistical probability.

Association and correlation do not necessarily imply causation, but they often provide an impetus for more study and confirmation.
**How can association be determined?**

This can be done by testing whether there is a numerical difference between the health indices (rates, means, and so forth) that describe the populations.

If there were no numerical differences between the health indices for the various populations and if we expected a difference, we might consider potential errors (bias) in method, sampling, data collection, and so forth, which may have masked a true difference in the populations studied. If such errors are likely to have happened, the study needs to be redesigned and repeated.

---

**Type I and type II errors**

Type II error: An association may exist in the population but, due to sampling variability, sample size, and the true difference in the parameters, the particular sample used may not reflect the true difference in the populations.

Thus, in doing a statistical test of significance, we would not reject the null hypothesis (because there is no numerical difference) and we would be making a type II error.

Type I error is made when we reject the null-hypothesis when it is really true.

---

**Establishing Practical Significance**

**What are five explanations which may make an artifactual association between a factor and a health outcome?**

1. Information bias.
2. Selection bias.
3. Failure to control for confounding variables.
4. Ecologic fallacy.
5. Sampling variability or chance.

---

**Information Bias**

Occurs when information obtained regarding exposure and disease outcome is incorrect

Bias in:

Abstracting records
Interviewing people
Recall bias
Reporting bias

---

**Selection Bias**

MacMahon study on pancreatic cancer and coffee drinking:
Controls have reduced intake of coffee, i.e. their coffee consumption is abnormally low.

Controls did not represent the general population level of coffee consumption

The result that cancer of pancreas is related to coffee consumption was due to selection bias

---

**Confounding**

To evaluate whether factor A is a cause of disease B one must first exclude factor X as confounding factor

Factor X is a confounder when:
1. Factor X is known risk factor for disease B
2. Factor X is associated with factor A, but not a result of factor A

Example: To evaluate whether coffee drinking is a cause of pancreatic cancer, first eliminate the role of smoking as a confounder.

1. Smoking is a known risk factor for pancreatic cancer
2. Smoking is associated with coffee drinking but is not result of coffee drinking
Ecological Fallacy

Ecological fallacy is applying the findings concerning group of people to individuals.

Example: Per capita dietary fat consumption association with carcinoma of the colon. Countries with high fat consumption show higher level of carcinoma of colon compared with low fat consumption countries. Applying this result to individuals is an ecological fallacy.

Applications of the causal criteria to Snow’s study:

1. Temporality of the association between the risks of drinking polluted water (Which came first) and the disease outcome (Which happened after).
2. Strength of the association. This has been ascertained through the estimation of the RR. The exposed persons to the polluted water were (8) times at greater risk of dying from cholera as those unexposed persons.
3. Experiment. Snow’s study in 1854 was a quasi-experiment i.e. “as if experiment”. It was a natural experiment.

At the time of Snow’s work, the cholera organism had not been discovered and there was little evidence to support the hypothesis that water was the vehicle of the disease. Thus it was difficult for other scientists to accept Snow’s data.

The End