Question 1: What makes the first study a case-control study?
The first case study is a case-control study because it goes from effect to cause (Friis, R. Sellers, T. 2009 pg 295) selecting cases and controls and then collecting data about exposures attributable to disease but occurred before the onset of the disease (past exposures). It collected information about exposure retrospectively. Primary data collected from the hospital was used for the study.

Question 2: What makes the second study a cohort study?
The second case study is a cohort study because it goes from cause to effect (Friis, R. Sellers, T. 2009 pg 294). Information about exposure was collected at baseline while, information on the outcome was collected prospectively.

Question 3: Why might hospitals have been chosen as the setting for this study?
Using the hospital as a setting for this study will ensure that “all true cases have an equal probability of entering the study and no false case enters the study” (Friis, R. Sellers, T. 2009 pg 264). All incident cases in London and vicinity over a 4-year period (April 1948 – February 1952) were identified and enrolled. Subjects involved in incident cases are likely to have a better recall of past exposures. Also, there is a likelihood that hospital controls will be more willing to participate.

Question 4: What other sources of cases and controls might have been used?
- A disease registry
- Medical records
- Reportable disease statistics
might have been used as another source of cases and control (Friis, R. Sellers, T. 2009 pg 212 & 264).

Question 5: What are the advantages of selecting controls from the same hospitals as cases?
Advantages of selecting controls from the same hospital as cases are
- This will ensure controls are ideal; cases will have the same characteristics as controls except the exposure of interest
- There will be no characteristic affecting observed outcomes (Friis, R. Sellers, T. 2009 pg 265).
Question 6: How representative of all persons with lung cancer are hospitalized patients with lung cancer?
Hospitalized patients with lung cancer are well representative of all persons with lung cancer (Friis, R. Sellers, T. 2009 pg 271) because all incident cases in London and vicinity over a 4-year period (April 1948 – February 1952) were identified and enrolled in the case-control study.

Question 7: How representative of the general population without lung cancer are hospitalized patients without lung cancer?
Hospitalized patients without lung cancer are well representative of the general population without lung cancer (Friis, R. Sellers, T. 2009 pg 271).

Question 8: How may these representativeness issues affect interpretation of the study's results?
In the population, the frequency of the disease is small this might result in sampling error (Friis, R. Sellers, T. 2009 pg 391).

Question 9: From this table, calculate the proportion of cases and controls who smoked.
Proportion smoked, cases: = number of cigarette smokers (cases)/ total number of people in study = 1,350/ 2,714 = 0.49742 = 0.5

Proportion smoked, controls: = number of cigarette smokers (controls)/ total number of people in study = 1,296/ 2,714 = 0.47752 = 0.5

Question 10: What do you infer from these proportions?
These proportions are equal to one another.

Question 11a: Calculate the odds of smoking among the cases.
The odds of smoking among the cases = (1,350 * 61)/ (1,296 * 7) = 82350/ 9072 = 9.0773 = 9

Question 11b: Calculate the odds of smoking among the controls.
The odds of smoking among the controls = (1,296 * 7)/ (1,350 * 61) 9072/ 82350 = 0.1101 = 0.1

Question 12: Calculate the ratio of these odds. How does this compare with the cross-product ratio?
The ratio of these odds = 9: 0.1 = 90
The cross product = 9 * 0.1 = 0.9
Comparing the ratio of these odds with the cross product ratio = 90: 0.9 = 100.

Question 13: What do you infer from the odds ratio about the relationship between smoking and lung cancer?
The odds ratio of 9, I can infer that smoking increases the risk of lung cancer by 9 times in the cases than in the controls. As a result, smoking is a risk factor for lung cancer.

Question 14: Compute the odds ratio by category of daily cigarette consumption, comparing each smoking category to nonsmokers.

Odds ratios by category
Daily number of cigarettes 1 – 14, odds ratio = (565*61)/ (7* 706) = 34465 / 4942 = 6.9 = 7

Daily number of cigarettes 15 – 24, odds ratio = (445*61)/ (408*7) = 27145/ 2856 = 9.5 = 10

Daily number of cigarettes 25+, odds ratio = (340*61)/ (182*7) = 20740/ 1274 = 16.3 = 16

Question 15: Interpret these results
Cigarette smoking is a risk factor for lung cancer and there is an increase in the risk of lung cancer with an increase in the number of cigarettes smoked daily. Smoking 1 – 14 cigarettes daily implies that the cases are seven times as likely as the controls to be exposed to lung cancer. Smoking 15 – 24 cigarettes daily implies that the cases are ten times as likely as the controls to be exposed to lung cancer. Smoking 25+ cigarettes daily implies that the cases are sixteen times as likely as the controls to be exposed to lung cancer.

Question 16: What are the other possible explanations for the apparent association?
Other possible explanations for the apparent association include;
- An increase in the number a cigarette smoked daily is directly proportional to an increase in the risk of lung cancer. If the odds ratio by category is plotted on a graph there will be a dose response relationship.
- There is a strong association with cigarette smoking (risk) and lung cancer (disease).
- The association between lung cancer and cigarette smoking has been observed to be consistent upon repetition
- Cigarette smoking is specifically associated with lung cancer
- Exposure to cigarette smoking precedes lung cancer (Friis, R. Sellers, T. 2009 pg 77).

Question 17: How might the response rate of 68% affect the study's results?
The response rate of 68% might affect the study’s result by
- Selection bias
In this cohort study, Physicians who lived in England and Wales in October 1951 were selected. These are gainfully employed and healthy individuals. Ill and unemployed persons were not represented in this study (Friis, R. Sellers, T. 2009 pg 392).

Question 18: Compute lung cancer mortality rates, rate ratios, and rate differences for each smoking category. What does each of these measures mean?

<table>
<thead>
<tr>
<th>Smoking category</th>
<th>Mortality rates per 1000 person-years</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-14</td>
<td>(22 /38,600) * 1000 = 0.5699 = 0.60</td>
</tr>
<tr>
<td>15-24</td>
<td>(54/ 38,900) * 1000 = 1.3881 = 1.40</td>
</tr>
<tr>
<td>25+</td>
<td>(57/ 25,100) * 1000 = 2.2709 = 2.30</td>
</tr>
<tr>
<td>All smokers</td>
<td>(133/ 102,600) * 1000 = 1.2962 = 1.30</td>
</tr>
<tr>
<td>Total</td>
<td>(136/ 145,400) * 1000 = 0.9353 = 0.94</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smoking category</th>
<th>Rate ratios of cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-14</td>
<td>0.60: 0.07 = 8.6 = 9</td>
</tr>
<tr>
<td>15-24</td>
<td>1.40: 0.07 = 20</td>
</tr>
<tr>
<td>25+</td>
<td>2.30: 0.07 = 32.9 = 33</td>
</tr>
<tr>
<td>All smokers</td>
<td>1.30: 0.07 = 18.6 = 19</td>
</tr>
<tr>
<td>Total</td>
<td>0.94: 0.07 = 13.4 = 13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smoking category</th>
<th>Rate differences per 1000 person-years</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-14</td>
<td>0.60 – 0.07 = 0.53</td>
</tr>
<tr>
<td>15-24</td>
<td>1.40 – 0.07 = 1.33</td>
</tr>
<tr>
<td>25+</td>
<td>2.30 – 0.07 = 2.23</td>
</tr>
<tr>
<td>All smokers</td>
<td>1.30 – 0.07 = 1.23</td>
</tr>
<tr>
<td>Total</td>
<td>0.94 – 0.07 = 0.87</td>
</tr>
</tbody>
</table>

Mortality rates per 1000 person-years in each smoking category means;
- Mortality rate of lung cancer per 1000 person-years in the category 1 – 14 is 0.60
- Mortality rate of lung cancer per 1000 person-years in the category 15 – 24 is 1.40
- Mortality rate of lung cancer per 1000 person-years in the category 25+ is 2.30
- Mortality rate of lung cancer per 1000 person-years in all smokers is 1.30
- Mortality rate of lung cancer per 1000 person-years in total is 0.94

Rate ratio of cases in each smoking category means:
- The risk of lung cancer in the exposure in the smoking category of 1 – 14 is associated with 8.6 the risk of disease.
- The risk of lung cancer in the exposure in the smoking category of 15 – 24 is associated with 9 times the risk of disease.
- The risk of lung cancer in the exposure in the smoking category of 1 – 14 is associated with 20 times the risk of disease.
- The risk of lung cancer in the exposure in the smoking category of 25+ is associated with 33 times the risk of disease.
- The risk of lung cancer in all smokers is associated with 19 times the risk of disease.
- The risk of lung cancer in total is associated with 13 times the risk of disease.

Rate differences per 1000 person-years in each smoking category means the difference in mortality rates in each smoking category and the mortality rates of non smokers.
- The mortality rate difference of lung cancer in the category 1- 14 is 0.53
- The mortality rate difference of lung cancer in the category 15- 24 is 1.33
- The mortality rate difference of lung cancer in the category 25+ is 2.23
- The mortality rate difference of lung cancer in all smokers is 1.23
- The mortality rate difference in total is 0.87

Question 19: What proportion of lung cancer deaths among all smokers can be attributed to smoking? What is this proportion called?
The proportion of lung cancer deaths among all smokers attributed to smoking = (number of lung cancer deaths attributed to smoking)/(total number of lung cancer deaths)*100 = (133)/(136)* 100 = 97.79% = 98%

This proportion is called the population etiologic fraction (Friis, R. Sellers, T. 2009 pg 368).

Question 20: If no one had smoked, how many deaths from lung cancer would have been averted?
If no one had ever smoked, 133 deaths from lung cancer would have been averted.

Question 21: Which cause of death has a stronger association with smoking? Why?
Lung cancer has a stronger association with smoking. The attributable risk percent among smokers is higher in lung cancer than in cardiovascular disease.

Question 22: Calculate the population attributable risk percent for lung cancer mortality and for cardiovascular disease mortality. How do they compare? How do they differ from the attributable risk percent?
Population attributable risk percent = (Incidence in entire population - Incidence in unexposed) / (Incidence in entire population) * 100

Population attributable risk percent for disease mortality = (mortality in entire population – mortality in unexposed)/ (mortality in entire population) * 100

Population attributable risk percent for lung cancer mortality = (0.94 – 0.07)/ (0.94) * 100 = 0.87/ 0.94 * 100 = 92.5% = 93%

Population attributable risk percent for cardiovascular disease mortality = (8.87 - 7.32)/ (8.87)* 100 = 1.55/ 8.87* 100 = 17.4% = 17%

The population attributable risk percent for lung cancer mortality is higher than the population attributable risk percent for cardiovascular disease mortality hence, 93% of deaths attributable to lung cancer could have been prevented in the entire population if smokers had not smoked while, 17% of deaths attributable to cardiovascular disease could have been prevented in the entire population if smokers had not smoked.

The attributable risk percent may be interpreted thus; 95% of deaths due to lung cancer among smokers that could have been prevented if the smokers had not smoked while, 93% of deaths attributable to lung cancer could have been prevented in the entire population if smokers had not smoked. The population attributable risk percent may be interpreted thus; 23% of deaths due to cardiovascular disease among smokers could have been prevented if the smokers had not smoked while, 17% of deaths attributable to cardiovascular disease could have been prevented in the entire population if smokers had not smoked.
Question 23: How many lung cancer deaths per 1,000 persons per year are attributable to smoking among the entire population? How many cardiovascular disease deaths?

Lung cancer deaths per 1,000 persons per year attributable to smoking among the entire population = \( \frac{\text{mortality in entire population} - \text{mortality in unexposed}}{\text{mortality in entire population}} \) = \( \frac{0.94 - 0.07}{0.94} = 0.93 \)

Cardiovascular disease deaths per 1,000 persons per year attributable to smoking among the entire population = \( \frac{\text{mortality in entire population} - \text{mortality in unexposed}}{\text{mortality in entire population}} \) = \( \frac{8.87 - 7.32}{8.87} = 0.17 \)

0.93 lung cancer deaths per 1000 persons per year are attributable to smoking among the entire population while, 0.17 cardiovascular disease deaths per 1000 persons per year are attributable to smoking among the entire population.

Question 24: What do these data imply for the practice of public health and preventive medicine?

For the practice of public health and preventive medicine these data implies that:

- Quitting smoking reduces the mortality rate of lung cancer.
- The number of years of quitting smoking up to 19 years is inversely proportional to the mortality rate of lung cancer. The mortality rate of quitting smoking and rate ratios by <5 years is 0.67 and 9.6, by 5-9 years is 0.49 and 7.0, by 10-19 years is 0.18 and 2.6, by 20+ years is 0.19 and 2.7 respectively.
- There is an inverse dose response relationship with lung cancer mortality and the number of years of quitting smoking for up to 19 years.
- Public health education on smoking cessation can reduce the risk of lung cancer and the number of years of smoking cessation is directly proportional to the risk reduction of lung cancer up to 19 years.

Question 25: Compare the results of the two studies. Comment on the similarities and differences in the computed measures of association.

Differences in the computed measures of association

Rate ratio, determines the risk of lung cancer associated with each category of smoking and provides an evidence of quantity of exposure of cigarette smoking by category in association.
with lung cancer. It studies the risk of lung cancer for each category of cigarette smoking;

- The risk of lung cancer in the category that smoked 1 – 14 cigarettes daily is 8.1 times the risk of disease.
- The risk of lung cancer in the category that smoked 15 – 24 cigarettes daily is 19.8 times the risk of disease.
- The risk of lung cancer in the category that smoked 25+ cigarettes daily is 32.4 times the risk of disease.
- The risk of lung cancer in all smokers is 18.5 times the risk of disease.

Odds ratio, it tells the odds of having an exposure to cigarette smoking in those with lung cancer in comparison with those who do not have lung cancer. It determines the risk of lung cancer for each category of cigarette smoking;

- The category of smokers that smoked 1 – 14 cigarettes daily are 7 times as likely as the controls to be exposed.
- The category of smokers that smoked 15 – 24 cigarettes daily are 9.5 times as likely as the controls to be exposed.
- The category of smokers that smoked 25+ cigarettes daily is 16.3 times as likely as the controls to be exposed.
- All smokers are 9.1 times as likely as the controls to be exposed

**Similarities in the computed measures of association**

- They are both used to in hypothesis testing of the risk of lung cancer associated with cigarette smoking
- The odds ratio and rate ratio for non smokers is 1. The rate ratio of 1 implies that, there is no risk of lung cancer among non smokers. The odds ratio of 1 implies that, that non-smoking is not a risk factor for lung cancer.

**Question 26:** What are the advantages and disadvantages of case-control vs. cohort studies?

**Advantages of case-control vs. cohort studies**

**Sample size;** it is smaller in size  
**Costs;** it is cost effective  
**Study time;** it saves time  
**Rare disease;** it is useful for studying rare diseases  
**Multiple exposure;** it can look at multiple exposures  
**Progression, spectrum of illness;** it is useful for studying diseases with long latency period between exposure and symptoms of disease
Advantages of cohort vs. case-control studies

**Multiple exposures**: it can be used in multiple exposures

**Rare exposure**: it is useful for studying rare exposures

**Diseases rates**: it is used in risk estimation, it allows direct determination of risk and it provides a stronger evidence of exposure-disease association.

**Multiple outcomes**: it allows examination of multiple outcomes

**Progression, spectrum of illness**: it provides evidence about lag time between exposure and disease

**Recall bias**: it minimizes recall bias

Disadvantages of case-control vs. cohort studies

**Recall bias**: it is prone to recall bias

**Rare exposure**: it is not useful for rare exposures

**Disease rates**: disease rates cannot be computed. It is unable to provide direct estimate of risk

**Multiple outcomes**: it assesses single outcome disease only

**Rare exposure**: it cannot be used for rare exposure

Disadvantages of cohort vs. case-control studies

**Cost**: it is expensive

**Sample size**: it can be extensive requiring large sample size

**Rare disease**: it is difficult to use with rare disease

**Study time**: it consumes time

**Selection bias**: it is prone to selection bias

**Loss of follow up**: loss to follow up can limit the sample size for analysis and can raise questions about results if loss is high

Question 27: Which type of study (cohort or case-control) would you have done first? Why? Why do a second study? Why do the other type of study?

I would have done a case-control study first. This was probably the first study on lung cancer and its association with cigarette smoking. A case-control study can be used when the etiology of chronic disease is not known. Also, it is cost effective, quick and easy to complete. It will show the association between lung cancer and cigarette smoking.

A second study is necessary to study cigarette smoking, a risk factor for lung cancer and estimate the risk of cigarette smoking in association with lung cancer.

A cohort study is needed because it allows a direct estimation of risk and provides a strong evidence of exposure-disease association. It tells the risk of developing lung cancer a disease based on exposure to cigarette smoking.
Question 28: Which of the following criteria for causality are met by the evidence presented from these two studies?

<table>
<thead>
<tr>
<th>Criteria</th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong association</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Consistency among studies</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Exposure precedes disease</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Dose-response effect</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Biologic plausibility</td>
<td>YES</td>
<td></td>
</tr>
</tbody>
</table>

References

Friis, R. Sellers, T. (2009) *Epidemiology for Public Health Practice* Jones and Bartlett Publishers Sudbury, Massachusetts