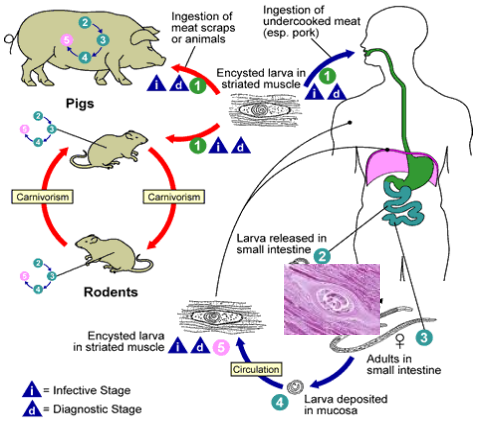
Design of studies

EXERCISES

María J. Vilar

Dep. of Food Hygiene and Environmental Health



28-31. 05. 2013

Exercise 1

An investigator wants to estimate the mean birth weight of infants to mothers who were tested positive to toxoplasma. The mean birth weight of infants from mothers tested negative is 3,510 grams with a standard deviation of 385 grams. How many women tested positive must be enrolled in the study to ensure that a 95% confidence interval estimate of the mean birth weight of their infants has a margin of error not exceeding 100 grams?

Exercise 2

An investigator wants to estimate the proportion of pigs at the biggest farm which currently eat a specific feed (i.e., the prevalence of eating that feed). How many pigs should be involved in the study to ensure that a 95% confidence interval estimate of the proportion of pigs which are feed with the specific feed is within 5% of the true proportion?

$$n = p(1-p)\left(\frac{Z}{E}\right)^2$$

Exercise 3a

An investigator wants to estimate the prevalence of Anisakis among women who ate cod living in a coast city. How many women must be involved in the study to ensure that the estimate is precise? National data suggest that 1 in 235 women are diagnosed with an infection by Anisakis. This translates to a proportion of 0.0043 (0.43%) or a prevalence of 43 per 10,000 women. Suppose the investigator wants the estimate to be within 10 per 10,000 women with 95% confidence.

Exercise 3b

An investigator wants to estimate the prevalence of Anisakis among women who ate cod living in a coast city. How many women must be involved in the study to ensure that the estimate is precise? National data suggest that 1 in 235 women are diagnosed with an infection by Anisakis. This translates to a proportion of 0.0043 (0.43%) or a prevalence of 43 per 10,000 women. Suppose the investigator wants the estimate to be within 10 per 10,000 women with 95% confidence.

A sample of size n=16448 will ensure that a 95% confidence interval estimate of the prevalence of Anisakis is within 0.10 of its true value. This is a situation where investigators might decide that a sample of this size is not feasible. Suppose that the investigators thought a sample of size 5000 would be reasonable from a practical point of view. How precisely can we estimate the prevalence with a sample of size n=5000?

Exercise 4

An investigator wants to plan a clinical trial to evaluate the efficacy of a new vaccine drug to protect against *E. coli*. The plan is to enroll participants and to randomly assign them to receive either the new drug or a placebo. Antibody levels will be measured in each participant after 4 weeks on the assigned treatment. Based on prior experience with similar trials, the investigator expects that 10% of all participants will be lost to follow up or will drop out of the study over 4weeks.

A 95% confidence interval will be estimated to quantify the difference in mean antibody levels between patients taking the new drug as compared to placebo. The investigator would like the margin of error to be no more than 3 units (error). How many patients should be recruited into the study? The standard deviation of antibody levels is 17.1.

$$n_i = 2\left(\frac{Z\sigma}{E}\right)^2$$

Exercise 5

An investigator hypothesizes that in people free of RF, a risk factor for infection by foodborne pathogen X, is higher in those who drink at least 2 glasses of non chlorinated water. A cross-sectional study is planned to assess the mean RF levels in people who drink at least two glasses of water per day. The mean RF level in people free of pathogen X is reported as 95.0 mg/dL with a standard deviation of 9.8 mg/dL.7 If the mean RF level in people who drink at least 2 glass of water per day is 100 mg/dL, this would be important clinically. How many patients should be enrolled in the study to ensure that the power of the test is 80% to detect this difference? A two sided test will be used with a 5% level of significance.

$$n = \left(\frac{Z_{1-\alpha/2} + Z_{1-\beta}}{ES}\right)^2 \qquad ES = \frac{|\mu_1 - \mu_0|}{\sigma}$$

Exercise 6

The investigators planned to randomly assign patients with recurrent *Clostridium difficile* infection to either antibiotic therapy or to duodenal infusion of donor feces. In order to estimate the sample size that would be needed, the investigators assumed that the feces infusion would be successful 90% of the time, and antibiotic therapy would be successful in 60% of cases. How many subjects will be needed in each group to ensure that the power of the study is 80% with a level of significance $\alpha = 0.05$?

Adapted from a randomized clinical trial reported in the New England Journal of Medicine (Jan. 2013)

Exercise 7

A researcher wants to estimate the prevalence of the pathogen X in conventional pig farms of whole Finland. The investigator plans on using a 95% confidence interval and wants an precision of 5%. The expected prevalence is unknown, but the investigators conduct a literature search find that the prevalence of pathogen X is 45%.

How many farms should be included in the study if prevalence were 45%? And how many in the region Pirkanmaan? And if the expected prevalence were 70%?

Would the sample increase or decrease if the error is higher, 10%?

Data in Excel file: Study design

Exercise 8

A risk ratio can be calculated from a case-control study, but the preferred measure of association is the odds ratio.

a. False

b. True

Exercise 9

What is the purpose of the control group in a case-control study?

a. To provide information on the disease distribution in the population that gave rise to the cases.

b. To provide information on the exposure distribution in the population that gave rise to the cases

Exercise 10

In 2004 there was an outbreak of Salmonellosis on the south of country Y. Over a period of a few weeks there were 20 cases reported to Y's health authorities, most of the infected persons were residents of city X. X's health department requested help in identifying the source from Y's health authorities. The investigators quickly performed descriptive epidemiology. The epidemic curve indicated a point source epidemic, and most of the cases lived in the X area, although some lived as far away. They conducted hypothesis-generating interviews, and taken together, the descriptive epidemiology suggested that the source was one of five or six food establishments in the X area, but it wasn't clear which one. Consequently, the investigators wanted to conduct an analytic study to determine which restaurant was the source.

What kind of study should the Y's health investigators do?

a. Retrospective cohort
b. Prospective cohort
c. Ecological study
d. Case-control study
e. Cross-sectional survey
f. Interventional (clinical trial)

Exercise 11

One of the hallmarks of analytical studies is that they compare groups to determine whether there appears to be an association

a. True

b. False

Exercise 12

Which of the following statements about case-control studies is true?

a. A retrospective cohort study is the same as a case control study.

b. Case control studies begin by identifying diseased subjects (cases) and a comparison group of non-diseased subjects (controls) in order to compare the exposure distribution between these two groups.

c. Case control studies begin by identifying non-diseased subjects (controls) and comparing the incidence of disease in the controls and diseased subjects (cases).

d. None of the above

Exercise 13

Intervention studies are most similar to which of the following study designs? (Select the ONE best answer.)

a. Retrospective Cohort Study
b. Prospective Cohort Study
c. Case Control Study.
d. None of the above

Exercise 14

In a retrospective cohort study the investigators "jump back in time" to identify and enroll subjects at a point in time before any of them had developed the outcome of interest

a. True

b. False

Exercise 15

Random error and systematic error are the same thing

a. True

b. False