Dawson-Saunders 1990

is-

2.5 SUMMARY

This chapter illustrates the study designs most frequently encountered in the medical literature. In medical research, either subjects are observed or experiments are undertaken. Experiments involving humans are called trials. Experimental studies may also use animals and tissue, although we did not discuss them as a separate category; the comments pertaining to clinical trials are relevant to animal and tissue studies as well.

Each type of study discussed has it advantages and disadvantages. Randomized, controlled clinical trials are the most powerful designs possible in medical research, but they are often expensive and time-consuming. Well-designed observational studies can provide useful insights on disease causation, even though they do not constitute proof of causes. Cohort studies are best for studying the natural progression of disease or risk factors for disease; case-control studies are much quicker and less expensive. Cross-sectional studies provide a snapshot of a disease or condition at one time, and we must be cautious in inferring disease progression from them. Case-series studies should be used only to raise questions for further research.

As much as possible, we have used Presenting Problems from later chapters to illustrate different study designs. We will point out salient features in the design of the Presenting Problems as we go along, and we will return to the topic of study design again after all the prerequisites for evaluating the quality of journal articles have been presented.

EXERCISES

Read the descriptions of the following studies and determine the study design used.

1. Kremer et al (1987) designed a study to determine the efficacy of fish oil dietary supplements in patients with rheumatoid arthritis. They were particularly interested in the effect of the fish oil on the inhibition of neutrophil leukotriene levels. The study involved a group of 40 patients with class I, II, or III rheumatoid arthritis; each patient was given either a dietary supplement or a placebo for 14 weeks, but the treatment assignment was not randomized. From weeks 14 to 18, all patients took a placebo for this four-week period; then they were given the opposite treatment (dietary supplement or placebo) from weeks 1 to 14 for the next 14 weeks.

2. A study by O'Malley and Fletcher (1987) looked at the efficacy of the breast self-examination (BSE) as a screening test for breast cancer by reviewing studies published on this topic. The authors found the sensitivity of BSE (the percentage of women with breast cancer who have a positive BSE) to be much lower than the sensitivity of a clinical breast examination or mammography. Although training increases the use of BSE and its sensitivity, the number of false-positives (women without breast cancer who have a positive BSE) also increases. The authors suggest the need for a controlled trial on BSE before advocating its use as a screening device.

3. Kilbourne et al (1983) investigated an epidemic in Spain involving multiple organ systems. Patients presented with cough, dyspnea, pleuritic chest pain, headache, fever, and bilateral pulmonary infiltrates. Although an infectious agent was first suspected, a strong association with food oil sold as olive oil but containing a high proportion of rapeseed oil was detected. Epidemiologic studies found that virtually all patients had ingested such oil but that unaffected persons had rarely done so.

4. Knutson et al (1981) treated wound, burn, and ulcer patients using granulated sugar combined with povidone-iodine. The study was undertaken from January 1976 to August 1980; during that time, 759 patients were treated. Of these, 154 were treated with the standard therapy and the remaining 605 were treated with sugar. Uniformity in treatment and judgment regarding the healing process were enhanced by using three physician-investigators to oversee the process and by documenting wound healing with 35-mm transparencies. The investigators reported that a much lower percentage of patients treated with the sugar and povidone-iodine mixture required skin grafts than those given the standard treatment; the therapy was painless, and changing the burn dressings was facilitated.

5. Colditz et al (1987) reported on the relationship between menopause and risk of coronary heart disease in women. Subjects in the study were selected from the Nurses' Health Study originally completed in 1976; the study included 120,000 married female registered nurses, aged 30–55. Colditz and his colleagues identified 116,000 of these women who were premenopausal or had a known type of menopause and did not have a diagnosis of coronary heart disease at the beginning of the study. The investigators were interested in determining whether the occurrence of menopause alters

the risk of coronary heart disease—specifically, whether the influence of menopausal status is altered by the use of postmenopausal estrogen. The original survey provided information on the subjects' age, parental history of myocardial infarction, smoking status, height, weight, use of oral contraceptives or postmenopausal hormones, and history of myocardial infarction or angina pectoris, diabetes, hypertension, or high serum calcium levels. Follow-up surveys were done in 1978, 1980, and 1982, and the data were 95.4% complete.

6. Bartle, Gupta, and Lazor (1986) designed a study to examine the association between nonsteroidal anti-inflammatory drug use and acute nonvariceal upper gastrointestinal tract bleeding. The association between consumption of acetylsalicylic acid and upper gastrointestinal bleeding is well established; however, no information was available on non-acetylsalicylic acid, nonsteroidal anti-inflammatory drugs. The medical records were reviewed to obtain medication histories of 57 consecutive patients with nonvariceal acute upper gastrointestinal tract hemorrhage presenting at a medical center, and 123 sex-matched and age-matched controls were in the study. (The process of sex and age matching ensures a control group that is similar to the cases with respect to gender and age.) The investigators found that a larger proportion of patients than of controls had taken nonsteroidal anti-inflammatory drugs.

the sum of squares by $(n - 1)$, not n, to obtain the best estimate of the variance, σ^2.

Exercise 6M

(Each branch is either true or false.)

1. The events A and B are mutually exclusive, so:

(a) Prob(A or B) = Prob(A) + Prob(B);
(b) Prob(A and B) = 0;
(c) Prob(A and B) = Prob(A) Prob(B);
(d) Prob(A) = Prob(B);
(e) Prob(A) + Prob(B) = 1.

2. The probability of a woman aged 50 having condition X is 0.20 and the probability of her having condition Y is 0.05. These probabilities are independent:

(a) The probability of her having both conditions is 0.01.
(b) The probability of her having both conditions is 0.25.
(c) The probability of her having either X, or Y, or both is 0.24.
(d) If she has condition X, the probability of her having Y also is 0.01.
(e) If she has condition X, the probability of her having Y also is 0.20.

3. The following variables follow a Binomial Distribution:

(a) number of sixes in 20 throws of a die;
(b) human weight;
(c) number of a random sample of patients who respond to a treatment;
(d) number of red cells in 1 ml of blood;
(e) proportion of hypertensives in a random sample of adult men.

4. If a coin is spun twice in succession:

(a) the expected number of tails is 1.5;
(b) the probability of two tails is 0.25;
(c) the number of tails follows a Binomial Distribution;
(d) the probability of at least one tail is 0.5;
(e) the distribution of the number of tails is symmetrical.

Exercise 7M

(Each branch is either true or false.)

1. The Normal Distribution:

(a) is also called the Gaussian Distribution;
(b) is followed by many variables;
(c) is so called because it is the one which is usually followed by naturally occurring quantities;
(d) is followed by all measurements made in healthy people;
(e) is the distribution towards which the Poisson Distribution tends as its mean increases.

2. The Standard Normal Distribution:

(a) is skew to the left;
(b) has mean = 1.0;
(c) has standard deviation = 0.0;
(d) has variance = 1.0;
(e) has the median equal to the mean.

3. The PEFRs of a group of 11-year-old girls are Normally distributed with mean 300 l/min and a standard deviation 20 l/min.

(a) About 95 per cent of the girls have PEFR between 260 and 340 l/min.
(b) 50 per cent of the girls have PEFR above 300 l/min.
(c) The girls have healthy lungs.
(d) About 5 per cent of girls have PEFR below 260 l/min.
(e) All the PEFRs must be less than 340 l/min.

4. The mean of a large sample:

(a) is always greater than the median;
(b) is calculated from the formula $\Sigma x_i/n$;
(c) is from an approximately Normal Distribution;
(d) increases as the sample size increases;
(e) is always greater than the standard deviation.

5. If X and Y are indep
Distributions, a Norm

(a) $5X$;
(b) X^2;
(c) $X + 5$;
(d) $X - Y$;
(e) X/Y.

Exercise 7E

In this exercise we shall ret
to decide how well they cc

1. From the box and wl
(see 4E solution if you hav
look like a Normal Distril

2. Construct a normal
ordered already in the ste
$i = 1$ to 40 and obtain th
from Table 7.1. Now pl
blood gucose.

3. Does the plot appe
Normal Distribution?

The SE of a mean

(b) is the accuracy with which each observation is measured;
(c) is a measure of how far the sample mean is likely to be from the population mean;
(d) is proportional to the number of observation;
(e) is greater than the estimated standard deviation of the population.

2. 95 per cent confidence limits for the mean estimated from a set of observations:

(a) are limits between which, in the long run, 95 per cent of observations fall;
(b) are a way of measuring the precision of the estimate of the mean;
(c) are limits within which the sample mean falls with probability 0.95;
(d) are limits which exclude the population mean with probability 0.05;
(e) are a way of measuring the variability of a set of observations.

3. If the size of a random sample were increased, we would expect:

(a) the mean to decrease;
(b) the standard error of the mean to decrease;
(c) the standard deviation to decrease;
(d) the sample variance to increase;
(e) the degrees of freedom for the estimated variance to increase.

4. The prevalence of a condition in a population is 0.1. If the prevalence is estimated repeatedly from samples of size 100, these estimates will form a distribution which:

(a) is a sampling distribution;
(b) will be approximately Normal;
(c) will have mean = 0.1;
(d) will have variance = 9;
(e) will be Binomial.

5. It is necessary to estimate the mean FEV1 by drawing a sample from a large population. The accuracy of the estimate will depend on:

(a) the mean FEV1 in the population;
(b) the number in the population;
(c) the number in the sample;
(d) the way the sample is sel
(e) the variance of FEV1 in

Exercise 8E

Table 8E.1 shows data from
subjects were all patients atte
month period.

Table 8E.1. Plasma m
treatment regimes (M

Treatment
Insulin
All non-insulin therapy
Oral hypoglycaemic the
Dietary restriction alon

N.B. Fifteen patients whos
both insulin and oral hypogly

1. Find the standard err
group.

2. Find the standard erro
between patients on oral h
restriction alone. Find a 95

3. Find the standard er
magnesium in insulin-treate
cent confidence interval for

4. What can be calculate
is determined by type or sev

5. How many patients w
plasma magnesium to withi

at least one significant result; we are more likely to get one than not. The expected number of spurious significant results is $20 \times 0.05 = 1$.

Many medical research studies are published with large numbers of significance tests. These are not usually independent, being carried out on the same set of subjects, so the above calculations do not apply exactly. However, it is clear that if we go on testing long enough we shall find something which is 'significant'. We must beware of attaching too much importance to a lone significant result among a mass of non-significant ones. It may be the one in twenty which we should get by chance alone.

This is particularly important when we find that a clinical trial or epidemiological study gives no significant difference overall, but does so in a particular subset of subjects, such as women aged over 60. A remarkable paper by Lee *et al.* (1980) demonstrates this. These authors simulated a clinical trial of the treatment of coronary artery disease by allocating 1073 patient records from past cases into two 'treatment' groups at random. They then analysed the outcome as if it were a genuine trial of two treatments. The analysis was quite detailed and thorough. As we would expect, it failed to show any significant difference in survival between those patients allocated to the two 'treatments'. Patients were then subdivided by two variables which affect prognosis, the number of diseased coronary vessels and whether the left ventricular contraction pattern was normal or abnormal. A significant difference in survival between the two 'treatment' groups was found in those patients with three diseased vessels (the maximum) and abnormal ventricular contraction. As this would be the subset of patients with the worst prognosis, the finding would be easy to account for by saying that the superior 'treatment' had its greatest advantage in the most severely ill patients! As the authors show, it is in fact explained by small chance differences in other prognosis indicators between the two 'treatment' groups in this subset. The moral of this story is that if there is no difference between the treatments overall, significant differences in subsets are to be treated with the utmost suspicion.

Exercise 9M

(Each branch is either true or false.)

1. **In a case-control study, patients with a given disease drank coffee more frequently than did controls, and the difference was highly significant. We can conclude that:**

(a) drinking coffee causes the disease;

(b) there is evidence of a
drinking in the sample
(c) the disease is not relate
(d) eliminating coffee wou
(e) coffee and the disease

2. **In a comparison of tw
had higher readings
readings on the mini-p
the difference between**

(a) the test statistic may be
meter;
(b) the null hypothesis is th
higher than the other;
(c) a one-tailed test of sig
(d) the test statistic shoul
$p = \frac{1}{2}$) if the null hypot
(e) the instruments sho

3. **When comparing the n**

(a) the null hypothesis is t
(b) the null hypothesis is t
(c) standard error of the
means;
(d) the standard errors of
(e) the test statistic is the

4. **In a small randomize
myocardial infarction,
the control group, b
conclude that:**

(a) the treatment is useles
(b) there is no point in co
(c) the reduction in mor
treatment immediatel
(d) we should keep addi
comparison of two pr
(e) we should carry out a

(b) there is evidence of a real relationship between the disease and coffee drinking in the sampled population;

(c) the disease is not related to coffee drinking;

(d) eliminating coffee would prevent the disease;

(e) coffee and the disease always go together.

2. In a comparison of two methods of measuring PEFR, 6 of 17 subjects had higher readings on the Wright peak flowmeter, 10 had higher readings on the mini-peak flowmeter and one had the same on both. If the difference between the instruments is tested using a sign test:

(a) the test statistic may be the number with the higher reading on the Wright meter;

(b) the null hypothesis is that there is no tendency for one instrument to read higher than the other;

(c) a one-tailed test of significance should be used;

(d) the test statistic should follow the Binomial Distribution ($n = 16$ and $p = \frac{1}{2}$) if the null hypothesis were true;

(e) the instruments should have been presented in random order.

3. When comparing the means of two large samples using the Normal test:

(a) the null hypothesis is that the sample means are equal;

(b) the null hypothesis is that the means are not significantly different;

(c) standard error of the difference is the sum of the standard errors of the means;

(d) the standard errors of the means must be equal;

(e) the test statistic is the ratio of the difference to its standard error.

4. In a small randomized double-blind trial of a new treatment in acute myocardial infarction, the mortality in the treated group was half that in the control group, but the difference was not significant. We can conclude that:

(a) the treatment is useless;

(b) there is no point in continuing to develop the treatment;

(c) the reduction in mortality is so great that we should introduce the treatment immediately;

(d) we should keep adding cases to the trial until the Normal test for comparison of two proportions is significant;

(e) we should carry out a new trial of much greater size.