

Clinical Biostatistics

Exercise: Significance Tests

In a double blind comparison of two ointments, containing calcipotriol or betamethasone, for the treatment of psoriasis, 345 subjects were given one ointment on the left side of the body and the other on the right, the side being chosen at random. The severity of the condition was assessed in terms of a score. The score was significantly lower ($P < 0.001$) for the calcipotriol side than the betamethasone side.

(a) What is meant by ' $P < 0.001$ '?

The sample size was calculated to allow detection of a difference of 5% between treatments in mean change in score, assuming a standard deviation for change in score of 25%, a type I error of 0.05 ($\alpha = 5\%$, two tailed) and a type II error of 0.10 (power = 90%) (*Lancet* 1991; **337**: 193-6).

(b) What is meant by type I error and type II error?

In a randomized trial of morphine vs placebo for the anaesthesia of mechanically ventilated pre-term babies, it was reported that morphine-treated babies showed a significant reduction in adrenaline concentrations during the first 24 hours (median change -0.4 nmol/L, $P < 0.001$), which was not seen in the placebo group (median change 0.2 nmol/L, $P < 0.79$) (*Lancet* 1993; **342**: 324-7).

(c) What is wrong with this approach to the analysis?

(d) Suggest a better method.

In a study of cancer and atmospheric nuclear weapons tests, exposed servicemen and civilian nuclear weapons workers were compared to a control group matched for age and occupation. One sided tests and 90% confidence intervals were used to compare the exposed group to the control group. The tests were done in the direction of the observed difference. The exposed groups were compared to the general population using standardized mortality ratios (SMRs). Two sided tests and 95% confidence intervals were used to compare the exposed group to the general population (*BMJ* 1993; **307**: 1530-5).

(e) What problems are there in this approach to testing?

(f) What two sided test would be equivalent to a one sided test at the 5% level in the direction of the difference?

In a study of treatments for menorrhagia during menstruation, 76 women were randomized to one of three drugs (*BMJ* 1996; **313**: 579-82). The effects of the drugs were measured within the subjects by comparing three control menstrual cycles and three treatment menstrual cycles in each woman. The women were given no treatment during the control cycles. In each subject the control cycles were the three cycles preceding the treatment cycles. The authors reported that patients treated with ethamsylate used the same number of sanitary towels as in the control cycles. A significant reduction in the number of sanitary towels used was found in patients treated with mefenamic acid ($P < 0.05$) and tranexamic acid ($P < 0.01$) comparing the control periods with the treatment periods. The table below shows some of the results.

Effect of three drugs on duration of bleeding and sanitary towel usage			
	Ethamsylate	Mefenamic acid	Tranexamic acid
<i>Mean duration (days) (SD)</i>			
Control	5.7 (1.1)	5.8 (1.3)	5.5 (1.4)
Treatment	5.7 (2.0)	5.3 (1.3)	4.9 (1.8)
<i>Mean No of sanitary towels (SD)</i>			
Control	25 (9.0)	25 (7.0)	23 (7.0)
Treatment	25 (9.0)	23 (9.0)	20 (6.0)

- (g) What problems are there in the design of this study?
- (h) What problems are there in the analysis?

In a study of bone density and falls in older women, 118 volunteers were randomized to receive either calcium supplements plus a program of exercise classes or to calcium alone for two years (*BMJ* 1997; **314**: 569).

- (i) The authors reported that they found no significant differences between the groups at baseline. Why were these tests of significance unnecessary?
- (j) The authors stated that ‘the difference between the groups in the number of women falling during the whole two year period was not significant ($P = 0.58$), but between 12 months and 18 months into the study the difference was significant ($P = 0.011$)’. Does the ‘ $P = 0.011$ ’ add anything of value to the results of this study?

Questions taken from Martin Bland and Janet Peacock: *Statistical Questions in Evidence-based Medicine*, Oxford University Press, Oxford, 2000.