Selected Exercises from IPS4e (M&M)

IPS4e Q 14.19

Exercise 7.37 (page 520) reports readings from 12 home radon detectors exposed to 105 picocuries per liter of radon:

```
 91.9 97.8 111.4 122.3 105.4  95.0
103.8 99.6  96.6 119.3 104.8 101.7
```

We wonder if the median reading differs significantly from the true value 105 (i.e. if a machine is just as likely to under- as to over-read)

(a) *Graph the data, and comment on skewness and outliers. A rank test is appropriate.*

(b) *We would like to test hypotheses about the median reading from home radon detectors:*

H_0: median = 105
H_a: median ≠ 105

To do this, apply the Wilcoxon signed rank statistic to the differences between the observations and 105. (This is the one sample version of the test.) What do you conclude?

(c) *[added by JH] What is the corresponding p-value if you use a simple sign test ?*

IPS4e Q 14.22

Exercise 12.11 presents the following data from a study of the loss of vitamin C in bread after baking:

Condition	Vitamin C (mg/100 g)	
Immediately after baking	47.62	49.79
One day after baking	40.45	43.46
Three days after baking	21.25	22.34
Five days after baking	13.18	11.65
Seven days after baking	8.51	8.13

The loss of vitamin C over time is clear, but with only 2 loaves of bread for each storage time we wonder if the differences among the groups are significant.

(**a**) *Use the Kruskal-Wallis test to assess significance, then write a brief summary of what the data show.*

(**b**) *Because there are only 2 observations per group, we suspect that the common chi-square approximation to the distribution of the Kruskal-Wallis statistic may not be accurate. The exact p-value (from the SAS software) is 0.0011. Compare this with your p-value from (a). Is the difference large enough to affect your conclusion?*

C**omment** by JH

The above analyses are a good example of the mistake made by the person who looks under the lamppost for his lost keys, even though he believes he lost them at a different place on the street-- just because there is more light under the lamppost! This is not a question of 5 "groups" or "conditions": the so-called groups have a very clear *time structure*, but the analysis used by M&M does not use this structure (If you interchange the rows (times) you still get the same p-value). Fortunately, the 'signal is strong enough here, and the noise from loaf to same-day loaf small enough that the differences are clear. A better -- more sensitive and focused -- analysis is to measure the trend (slope of regression line) in vitamin C over time -- just as your eye does! It is *not a question of whether* bread loses vitamin C, but *how quickly* it does. If the question were "after 1 day, is the loss more with certain of 5 *types* of bread than others, then a (*global*) Kruskal-Wallis or other statistic might be more appropriate-- but again, it is probably not a question of whether, but of how much.

[The same comment applies whether we are parametric or non-parametric. And JH will be returning to this issue when we study chi-square tests for proportions rather than means.]

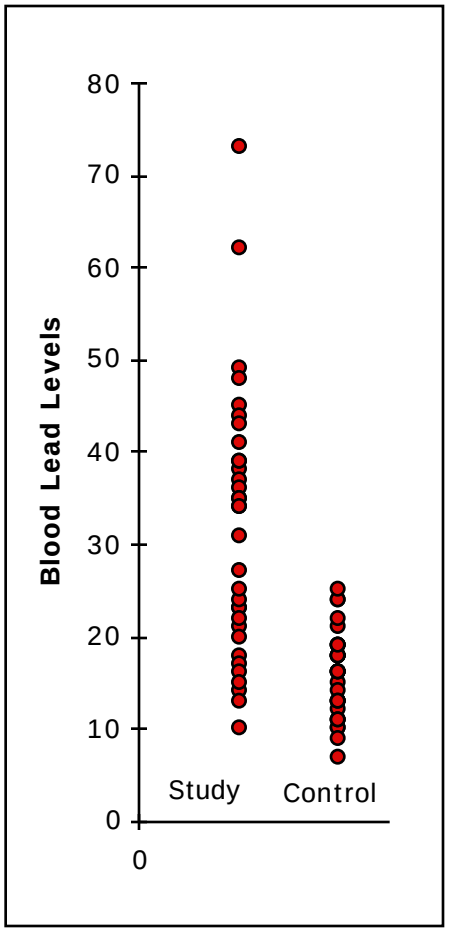
Homegrown Exercises, Ch 15

1 **LEAD ABSORPTION IN CHILDREN OF EMPLOYEES IN A LEAD-RELATED INDUSTRY.** Morton DE American J Epi 1982; 115:549-55.

Children can be exposed to lead from a variety of environmental sources. It has been repeatedly reported that children of employees in a lead-related industry are at increased risk of lead absorption because of the high levels of lead found in the household dust of these workers. A case-control study[a] was done in Oklahoma in 1978 to determine whether children of employees in a battery manufacturing plant had a higher prevalence of high levels of blood lead than children whose parents were not employed in a lead-related industry. The data obtained indicated that the blood lead levels of the study children were significantly greater than those of the control children.[b] None of the control children had blood levels of >30 µg/dl, while 53% of the exposed children had blood levels of >30 µg/dl. Trends indicated that the children whose fathers had higher lead exposure at work also had higher blood lead levels. However, the study children whose fathers had good personal hygiene had blood levels comparable to the control children. It appeared that only good personal hygiene, i.e., showering, shampooing and changing clothes and shoes before leaving work, was effective for lead containment. The mere changing of clothes and shoes appeared to be inadequate for lead containment.

Table 1 Blood levels of study and control children and the difference between blood lead levels of the two groups by matched pairs

	Lead levels				Lead levels		
pair	Study	Ctl	Diff	pair	Study	Ctl	Diff
1	38	16	22	18	10	13	-3
2	23	18	5	19	45	9	36
3	41	18	23	20	39	14	25
4	18	24	-6	21	22	21	1
5	37	19	18	22	35	19	16
6	36	11	25	23	49	7	42
7	23	10	13	24	48	18	30
8	62	15	47	25	44	19	25
9	31	16	15	26	35	12	23
10	34	18	16	27	43	11	32
11	24	18	6	28	39	22	17
12	14	13	1	29	34	25	9
13	21	19	2	30	13	16	-3
14	17	10	7	31	73	13	60
15	16	16	0	32	25	11	14
16	20	16	4	33	27	13	14
17	15	24	-9				



Questions

a *[if in one of the Epidemiology programmes] Do you agree that this should be called a case-control study? Why, or why not?*

b *STATE the alternative hypothesis, explicitly in terms of "location" and shapes of distributions. LIST the tests could one use to test it, if all one had was the data in the **Figure**. Do not carry out the tests, but LAY OUT THE STEPS your research assistant would need to follow to perform the test you consider most appropriate.*

c *Repeat **b** but now assuming one had the data in **Table 1**.*

2 [from A&B] **Obstetric records of (the mothers of) a group of children who died "suddenly and unexpectedly' (SUD) were compared with those of a group of live 'control' children**. Observations on the duration of the 2nd stage of labour were as follows:

```
S.U.D.   60, 25,  6, 8,  5, <5,   10,  25, 15, 10
Controls 13, 20, 15, 7, 75, 120*, 10, 100,  9, 25, 30
          *: terminated by surgical intervention.
```

a *[for students in one of the Epidemiology programmes] Do you agree that this should be called a case-control study? Why, or why not?*

b *Compute and compare the median duration of labour in each group and evaluate the statistical significance of the difference.*

3 **Clustering of teenage suicides after television news stories about suicide** DP Phillips, and LL Carstensen. NEJM Volume 315:685-689 September 11, 1986 Number 11

Abstract: We examined the relation between 38 nationally televised news or feature stories about suicide from 1973 to 1979 and the fluctuation of the rate of suicide among American teenagers before and after these stories. The observed number of suicides by teenagers from zero to seven days after these broadcasts (1666) was significantly greater than the number expected (1555; P = 0.008). The more networks that carried a story about suicide, the greater was the increase in suicides thereafter (P = 0.0004). These findings persisted after correction for the effects of the day of the week, the month, holidays, and yearly trends. Teenage suicides increased more than adult suicides after stories about suicide (6.87 vs. 0.45 percent). Suicides increased as much after general-information or feature stories about suicide as after news stories about a particular suicide. Six alternative explanations of these findings were assessed, including the possibility that the results were due to misclassification or were statistical artifacts. We conclude that the best available explanation is that television stories about suicide trigger additional suicides, perhaps because of imitation.

The 38 observations below are from Table 1 in the above article. In the beginning of the Results section, the authors state: ".. teenage suicides rose by 2.91 per story during the period 0-7 days after each story (p=0.008 by the 2-tailed Wilcoxon test)

Questions...

a *LIST suitable statistical tests for this comparison.*

b *HOW WOULD one verify the p-value the authors obtained?*

c *[Optional] Compare the total number observed with the total number expected, using a X^2 test for data in a "1×1" table [This is equivalent to using the Poisson distribution]*

Table 1: Fluctuations in the number of suicides in US teenagers zero to seven days after stories about suicide on network television news programs

o	exp	diff	o	exp	diff	o	exp	diff
34	38.01	-4.01	39	40.46	-1.46	45	45.3	-0.3
36	39.86	-3.86	40	38.43	1.57	40	45.14	-5.14
34	30.32	3.68	38	36.91	1.09	38	40.05	-2.05
28	29.77	-1.77	46	48.07	-2.07	45	40.99	4.01
40	33.83	6.17	58	48.5	9.5	43	41.93	1.07
30	31.9	-1.9	59	48.75	10.25	52	45.11	6.89
35	35.44	-0.44	57	49.67	7.33	48	43.78	4.22
40	37.95	2.05	56	48.59	7.41	41	37.99	3.01
46	41.84	4.16	54	45.61	8.39	61	40.82	20.18
43	43.85	-0.85	40	40.81	-0.81	33	43.71	-10.7
48	37.93	10.07	43	40.02	2.98	58	43.06	14.94
35	33.56	1.44	44	40.44	3.56	47	42.31	4.69
50	40.49	9.51	42	44.15	-2.15			

o: observed number of suicides

exp: expected number based on data for corresponding weeks in other years

diff: difference o – exp

4 Acetazolamide in prevention of acute mountain sickness: a double blind controlled crossover study

(article attached at end of this document, i.e., after Q5)

>>> Parts a-c refer only to Kilimanjaro portion of the expedition.

a "Those taking acetazolamide reached a higher altitude (11 versus 4 reached the summit)" (abstract).

[treating the outcome as binary] Carry out a statistical test to evaluate the 11 vs. 4 "successes in reaching the summit" (i) using the pairing (ii) ignoring the pairing.

[NB: for (i), the information provided is not sufficient, so do the test with each of the possible configurations]

b *[treating the outcome as ordinal]*

"Fig. 2 compares the altitudes reached by subjects taking the drug and those taking the placebo... the drug group showed a striking advantage (Wilcoxon signed rank sum test $p < 0.01$)" (last paragraph, 2nd page)

Presumably, they carried out a "Wilcoxon signed rank test" for paired data. They call it a "signed rank sum test" ... they used the terminology in A&B's textbook rather than in Bradford Hill's or M&Ms'. It would be better if publications called it the "Wilcoxon test for paired data".

Again, there is a small ambiguity, from the data in the Figure, as to what the 12 pairings of 'altitudes reached' were.

From the p-value reported, try to determine what the configuration i.e. the pairings must have been

NB: Do not be upset if you cannot replicate the p-value exactly, since it is not clear whether the authors' p-value is 1- or 2-sided, or how they handled ties, or whether they used exact methods (a table, or an enumeration, as I show in the diagonals of the table I worked out and put as a separate item "Wilcoxon Signed Rank Test: by Enumeration" on the web page), or by a Gaussian approximation.

c "In every pair the partner on acetazolamide had the lower symptom score." (first sentence of third page)

i *What value of the Wilcoxon signed rank statistic does this imply? (If you like, use the same shortcut used by Gauss*)*

ii *What other non-parametric test is suggested by this statement?*

d Kilimanjaro vs. Mt Kenya:

We could make two contrasts:

- self paired: using the data from the "self paired" crossover: i.e. use the data from the two expeditions; compare each person's data from the expedition on which (s)he was taking active treatment with the same person's data from the expedition on which (s)he was taking placebo... a one-sample test (within-person comparisons, a paired t-test with 23 df if we were using parametric tests)
- partner pairs: using the data from the "matched pairs": use only the data from the Mt. Kilimanjaro expedition; compare each treated person's data with his/her partner's data... again a paired t-test but with only 11 df, and between-person comparisons)

Although contrast (i) looks more powerful statistically (and is the one implied in the title of the paper), why is it the scientifically weaker one of the two in this study?

* There is a story about Karl Friedrich Gauss when he was in elementary school. His teacher got angry at the class and told them to add the numbers 1 to 100 and give him the answer by the end of the class. About 30 seconds later Gauss gave him the answer. The other children were adding the numbers like this: $1 + 2 + 3 + \ldots + 99 + 100 = ?$
But Gauss rearranged the numbers to add them like this:
$(1 + 100) + (2 + 99) + (3 + 98) + \ldots + (50 + 51) = ?$
Every pair of numbers adds up to 101. There are 50 pairs of numbers, so the answer is $50 \times 101 = 5050$.

5 Do Dogs Resemble Their Owners?

Roy MM & Christenfeld NJS. PSYCHOLOGICAL SCIENCE, Vol 15, No. 5, pp361-363, 2004 [full article available under Resources Ch 14]

ABSTRACT—We examined whether the frequent casual reports of people resembling their pets are accurate by having observers attempt to match dogs with their owners. We further explored whether any ability of observers to make such matches is due to people selecting dogs who resemble them, in which case the resemblance should be greater for predictable purebreds than for nonpurebreds, or is due to convergence, in which case the resemblance should grow with duration of ownership. Forty-five dogs and their owners were photographed separately, and judges were shown one owner, that owner's dog, and one other dog, with the task of picking out the true match. The results were consistent with a selection account: Observers were able to match only purebred dogs with their owners, and there was no relation between the ability to pair a person with his or her pet and the time they had cohabited. The ability to match people and pets did not seem to rely on any simple trait matching (e.g., size or hairiness). The results suggest that when people pick a pet, they seek one that, at some level, resembles them, and when they get a purebred, they get what they want.

{from METHODS} Each set of 15 pictures was viewed by 28 naive undergraduate judges who were participating for course credit. We constructed triads of pictures, each consisting of one owner, that owner's dog, and one other dog photographed at the same park. Each judge was shown the 15 owners from one dog park, one at a time, and instructed to identify which of the two possible dogs belonged to each person. In a Latin square design, each of the 14 incorrect dogs served as a foil for each dog, with the order of presentation randomized. Thus, within each set of 15 pairs, each owner-and-dog pair was presented with every other dog photographed at that park. A dog was regarded as resembling its owner if a majority of judges (i.e., more than 14) matched the pair.

RESULTS There was no evidence of any resemblance between nonpurebreds and their owners; of the 20 such dogs, there were 7 matches, 4 ties, and 9 misses, Chi_square[2df, N=20) = 0.64, n.s.. However, purebreds could be matched with their owners; of the 25 purebreds, there were 16 matches, 0 ties, and 9 misses, Chi_square[2df, N=25) = 6.75, p <.05. The difference between the matchability of nonpurebreds and purebreds was significant, Chi_square[2df, N=45) = 7.03, p <.05. ...

No. of judges/28 who correctly matched owner and dog
[Data supplied by author to JH; "mutt" = not purebred]

Mutt #	No./28	Purebred #	No./28
1	15	1	17
2	23	2	21
3	15	3	17
4	12	4	12
5	14	5	12
6	14	6	15
7	9	7	17
8	12	8	21
9	21	9	10
10	9	10	20
11	7	11	17
12	11	12	8
13	14	13	18
14	12	14	8
15	20	15	17
16	14	16	18
17	16	17	18
18	12	18	11
19	17	19	18
20	13	20	9
		21	16
		22	20
		23	11
		24	15
		25	10
Mean	14.00		15.04
SD	4.01		4.15

Questions on 5 (Do Dogs Resemble Their Owners?)

a *If a judge simply guessed (or tossed a coin) to decide which of the two dogs belonged to the owner, what is the probability that the judge's guess would be correct?*

b *If 28 judges guessed, what is the probability that (i) a majority i.e., more than 14 (what the authors call a 'match') (ii) exactly 14 (what the authors call a 'tie') and (iii) fewer than 14 (what the authors called a 'miss') would match the owner and the dog?* The Excel spreadsheet "Binomial Distributions (how shapes varies with n and p) in the Resources for Chapter 5 of course 607 can help you here.

c *If 28 judges guessed about 20 different owners (and their non-purebred dogs), for how many of the 20 would you expect there to be a match? a tie? a miss?*

d *How do these expected numbers compare with the numbers observed in the study (first paragraph of Results)?*

e *If, instead of the 3 categories the authors used, you used a simple dichotomy ">14" versus "14 or fewer" (i.e. a 'match' versus 'ties or miss', and if indeed judges were simply guessing, what is the probability of observing (i) 1 match in 1 owner (and its non-purebred) dog (ii) 7 or more matches in all 20?*

f *Repeat above calculations, but for the 25 purebred dogs.*

If instead of transforming the number, out of the 28 judgments, that were correct, into 2 or 3 categories, suppose you used you used the number/28 'as is' i.e., as a *number* between 0 and 28 (as in the Table above).

g *Under the null hypothesis that the judges were simply guessing, what is (i) the mean (i.e. expected) value (ii) variance and (iii) standard deviation of this number?*

h *What does the (alternative) hypothesis that the authors wished to test say about the expected value?*

i *Across the 25 purebreds, what is the average number of judges who correctly matched owner and dog?*

j *Using **g-i,** calculate a test statistic, and its associated (one-sided) p-value. Comment.*

In the majority of applications involving tests of means, one must estimate the variance or standard deviation from the data, and use the (wider) t-distribution to account for the extra uncertainty; here, in this example, under the null hypothesis, you know the variance. Thus , if under the null, the observations would have a Gaussian distribution, one could use the Z-distribution as the reference distribution.

k *In this situation, under the null hypothesis, is it reasonable to assume that the numbers would have a close-to-Gaussian distribution around the mean of 14?*

Imagine that the investigators had designed a more difficult test, where instead of *one* other dog, they had six other dogs.

l *Under this scheme, what would the expected (mean) value and the standard deviation of the number of judges who picked out the correct dog?*

m *Under this scheme, Why would the number not have a close-to-Gaussian distribution around this mean?*

n *Would one still be justified in using the Z-test to test whether across the 25 purebreds, the average number of judges who got it right was significantly higher than expected under the null? What if there were only 5 purebreds? what if there were 100?*

o Suppose you boss/chief (or the editor/referee for the journal) had never heard of the Central Limit theorem, is not convinced by your answers, and suggests that you perform a non-parametric test of the same null hypothesis for the 25 purebred dogs.

- i *List 2 such tests* (we have already used a variation on one of these, without giving it a formal name), *and indicate which should be the stronger (more powerful/sensitive) of the 2*
- ii *State the null (and alternative) hypothesis they test.*
- iii *Carry out the two tests and comment on the findings.*

p The authors compared the classifications of the 20 non-purebreds with those of the 25 purebreds, since their theory predicted that the accuracy with the purebreds should be better.

- i *Use the 3 different data-reduction methods (trichotomy, dichotomy, raw number correct) to compare the accuracy in the 20 versus the 25.*
- ii *Give two reasons why, for this type of situation, their chi-square methods should not be as sensitive as those based on the non-categorized numbers (Hint: one has to do with the 'granularity' of the data, the other with which tests do/do not take account of the directionality in the alternative hypothesis)*

CLINICAL RESEARCH

Acetazolamide in prevention of acute mountain sickness: a double-blind controlled cross-over study

M K GREENE, A M KERR, I B McINTOSH, R J PRESCOTT

Abstract

Twenty-four amateur climbers took part in a double-blind controlled cross-over trial of acetazolamide versus placebo for the prevention of acute mountain sickness. They climbed Kilimanjaro (5895 m) and Mt Kenya (5186 m) in three weeks with five rest days between ascents. The severity of acute mountain sickness was gauged by a score derived from symptoms recorded daily by each subject. On Kilimanjaro those taking acetazolamide reached a higher altitude (11 v 4 reached the summit) and had a lower symptom score than those taking placebo (mean 4·8 v 14·3). Those who had taken acetazolamide on Kilimanjaro maintained their low symptom scores while taking placebo on Mt Kenya (mean score 1·9), whereas those who had taken placebo on Kilimanjaro experienced a pronounced improvement when they took acetazolamide on Mt Kenya (mean score 2·5). Acute mountain sickness prevented one subject from completing either ascent. Acetazolamide was acceptable to 23 of the 24 subjects.

Acetazolamide is recommended as an acceptable and effective prophylactic for acute mountain sickness.

Introduction

The present popularity of short trekking holidays with rapid ascents has concentrated attention on the problem of acute mountain sickness. This is a symptom complex in which weakness, breathlessness, dizziness, and nausea impair performance and enjoyment for those who are unacclimatised and who venture over 3000 metres. It is usually mild and transient and at its worst within three days; in a few people, however, it progresses rapidly to life-threatening pulmonary or cerebral oedema. Incidence is highest when ascent is rapid and exertion great. Susceptibility has been reported greatest in the young, decreasing up to the age of 40.[1] At high altitudes, the hyperventilation which would compensate for a falling arterial oxygen is inhibited by the respiratory alkalosis which it induces. This happens chiefly during sleep and it is at that time that acetazolamide may have a beneficial effect by producing a metabolic acidosis.[2,3] The place of acetazolamide in the prevention of altitude sickness has been explored through decompression experiments and in several clinical trials. No trekking expedition, however, has met all the requirements of a double-blind controlled trial, and a cross-over trial has not been attempted.[3,4-8] One study[9] compared the performances of five climbers on two expeditions a year apart.

We set out to meet the necessary conditions for a trial of acetazolamide during an attempt on Africa's two highest peaks by the Scout East Africa Expedition in the summer of 1980.

Methods

The party comprised 24 British residents, including two women. None were professional sportsmen; five were medically trained. They were paired for age, sex, and likely activities, and each member of each pair was allocated at random to one of two treatment groups. Treatment group 1 received acetazolamide (sustained release 500 mg nightly) on five nights before and during the first ascent and identically presented placebo before and during the second ascent. The treatment order was reversed for group 2 (see fig 1, which also gives expedition objectives).

An error in allocation of capsules led to the re-arrangement of two pairs; the pairs of subject 22 and subject 20 and of subject 19 and subject 21 being changed to 19/20 and 21/22: the table shows the pairs actually used. The conclusions from statistical analysis are unaffected if these pairs are excluded, and they are therefore included.

The entire party travelled by air from London to Nairobi, by road to Kilimanjaro, and then on foot from 2000 m. Kit weighing about 10 kg was carried. Nights were spent at the hut camps provided.

Charing Cross Hospital, London W6 8RF
M K GREENE, medical student

Royal Hospital for Sick Children, Yorkhill, Glasgow G3 8SJ
A M KERR, DCH, MRCP, registrar in paediatrics (present appointment: senior registrar)

Stirling Royal Infirmary, Stirling FK8 2AU
I B McINTOSH, MB, CHB, clinical assistant in geriatrics and general practitioner

Edinburgh Medical School, Edinburgh
R J PRESCOTT, MSC, PHD, senior lecturer in medical statistics

On Mt Kenya the expedition divided into four activity groups, the participants being decided during the trek. The table gives identification numbers of subjects. Subjects 2, 3, 7, and 21 left in advance and climbed the central rock pinnacles Nelian (5174 m) and Batian (5186 m); subjects 4, 8, 12, and 19, acting as support party, climbed on rock and glacier around 5000 m; and subjects 6, 16, 17, and 23 circumnavigated the central pinnacles and climbed point Lenana (4972 m). The remainder climbed point Lenana only. The party slept in tents and carried an average of 10 kg each.

Before departure the medical project was explained and full co-operation obtained. Every subject recorded each day's symptoms nightly on a card. A list of common symptoms of acute mountain sickness was printed down one side of the squared card and expedition days across the top. One tick was to be entered if a symptom was experienced on a given day. Spaces were provided for: distance walked, load carried, night altitude, metres climbed, symptoms not related to acute mountain sickness, and medication. Comments were invited. Permitted drugs for treatment were: aspirin for headache; diphenoxylate and atropine (Lomotil) for diarrhoea; temazepam and nitrazepam for sleeplessness; and chloroquine, pyrimethamine, and dapsone for malarial prophylaxis.

Scores for acute mountain sickness were calculated from symptom cards by giving one point for mild headache, loss of appetite, feeling sick, severe inappropriate weakness, dizziness, depression, irritability, drowsiness, cough, and shortness of breath walking on the flat, and three points each for severe headache, vomiting, staggering, shortness of breath at rest, and frothy spit. If severe headache or shortness of breath at rest was scored then its milder form was not. Each subject's score was calculated for days spent off the mountains and was averaged to give a baseline score. This was subtracted from each ascent day's score, giving rise to some negative scores. The analysis uses three-day (ascent period) scores for each person, as both acute mountain sickness and acetazolamide were expected to have most effect during these periods. Each analysis started on the first night over 3000 m and so included the summit for everyone on Kilimanjaro and comparable altitudes for everyone on Mt Kenya (see fig 1).

The data do not conform to normal distributions, therefore non-parametric tests of significance (Wilcoxon and Spearman rank correlation tests) were used. The test used to examine interaction between the first and second treatment periods and ascents is based on that described by Hills and Armitage[9] with an adaptation to allow for the pairing in our design.

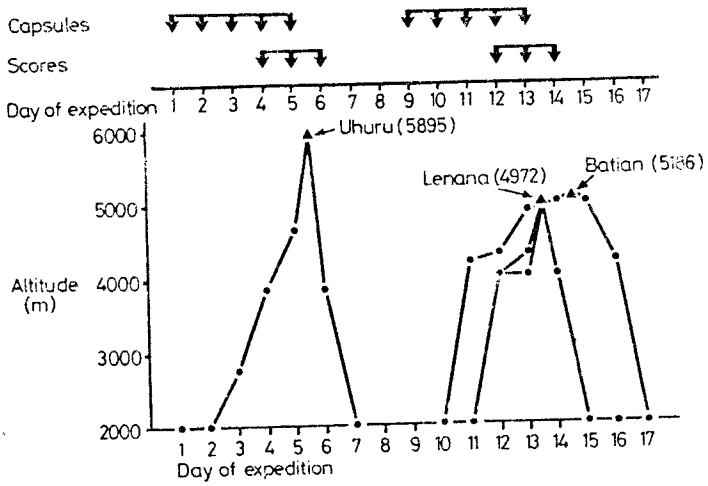


FIG 1—Expedition programme.

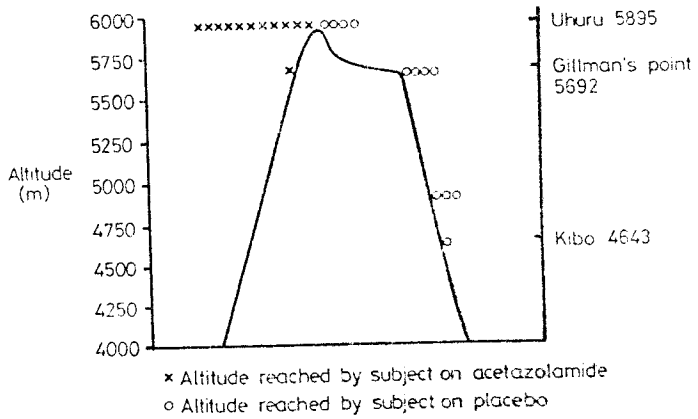


FIG 2—Altitudes reached on Kilimanjaro.

Symptom scores for each subject on each mountain in order of ascent (after deduction for non-ascent days symptoms)

Treatment group 1				Treatment group 2			
Subject No	Age	Acetazolamide Kilimanjaro	Placebo Mt Kenya	Subject No	Age	Placebo Kilimanjaro	Acetazolamide Mt Kenya
1	20	7	0	2	20	25	−1
3	20	13	7	4	21	19	5
5	49	3	3	6	43	17	9
7	36	4	missing	8	23	7	1
9	17	5	−1	10	18	9	3
11	45	6	−1	12	45	12	2
13	45	0	0	14	50	18	2
15	36	1	0	16	36	12	0
17	45	3	0	18	41	5	4
19	41	5	2	20	19	12	−1
21	27	9	9	22	19	18	−2
23	42 (F)	2	2	24	24 (F)	17	8
Total score		58	21			171	30
Mean		4·8	1·9			14·3	2·5

Results

No one left the expedition. Illness other than acute mountain sickness was trivial. Diarrhoea was reported on 11 days (five on acetazolamide and six others). Acute mountain sickness manifest as severe headache and vomiting prevented subject 24 from making the final six-hour ascent (on placebo) of the summit of Kilimanjaro. On Mt Kenya breathlessness at rest and severe inappropriate weakness prevented her from climbing (on acetazolamide) above 3500 m. On this occasion her two tent mates (subjects 20 and 22) remained with her voluntarily; both were taking acetazolamide. As exclusion of these two subjects from statistical analysis does not alter the conclusions, they have been included. Compliance was in general excellent. On Mt Kenya subject 18 stopped his acetazolamide after taking one, saying that he felt ill, and subject 10 mislaid his acetazolamide after taking two. Subject 7 failed to complete his card on Mt Kenya. The double blind was entirely successful.

Of side effects, tingling in the extremities was reported by 7 out of 24 subjects (29%) on acetazolamide and three out of 24 (12%) on placebo. Diuresis was not remarked on, perhaps because of concurrent changes in living conditions. Nausea was reported on the first day of taking acetazolamide by two of 24 (8%) (none on placebo).

With the exceptions mentioned above, planned objectives were reached by all subjects on Mt Kenya.

Fig 2 compares the altitudes reached on Kilimanjaro by subjects on acetazolamide and placebo. Those taking acetazolamide showed a striking advantage (Wilcoxon signed rank sum test $p<0.01$). The symptom scores of each treatment group on each mountain also show an impressive advantage for those taking acetazolamide on Kilimanjaro

(see table). In every pair the partner on acetazolamide had the lower symptom score (Wilcoxon signed rank sum test $p < 0.001$). On Mt Kenya the two groups performed equally well. Those previously on placebo improved greatly on acetazolamide, while those previously on acetazolamide maintained their low scores although taking placebo. It is clear that the first treatment and ascent period had an effect on the second. The presence of this interaction was confirmed statistically using a test based on that described by Armitage and Hills ($p < 0.01$).[9]

Fig 3 shows the total cross-over experience graphically. Pairs 9/10 and 17/18 failed to take the full course of acetazolamide. No statistically significant association was found between the ages of subjects and their scores. On Kilimanjaro there was a suggestion of a negative association among subjects on acetazolamide (Spearman rank correlation coefficient $= -0.55$ ($0.05 < p < 0.1$)). The corresponding subjects on placebo, however, showed no correlation ($r = -0.04$), and the scores on Mt Kenya showed no correlation with age.

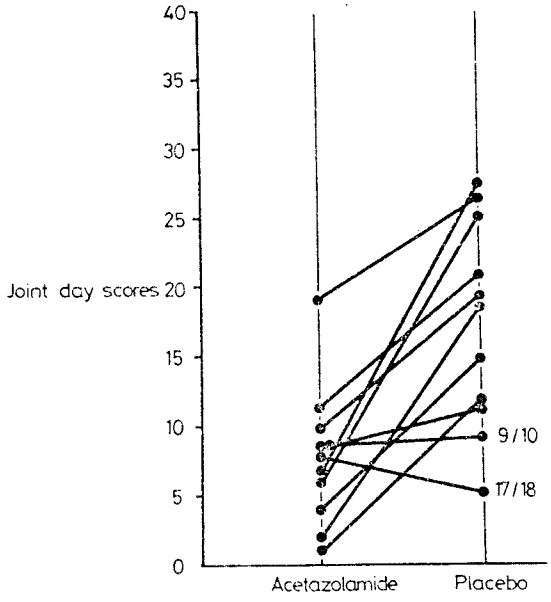


FIG 3—Joint symptom scores for each pair on each ascent. Numbers indicate pairs who failed to take full course of acetazolamide.

Discussion

The weaknesses of symptom reporting as a means of assessing acute mountain sickness are obvious but must be accepted, since there are no reliable signs of the mild form of acute mountain sickness. Advantages of our method were its simplicity and the promptness of reporting. Baseline scores helped to reduce the effect of personal differences in symptom recognition. Co-operation was excellent. The allocation of points for symptoms, while arbitrary, distinguishes clearly between the mild ones and those which indicate cerebral or pulmonary oedema.

Acetazolamide was found to be a useful prophylactic for acute mountain sickness on this expedition during rapid ascent on foot from 2500 m to 5000 m. The advantage to those on acetazolamide was most evident on the ascent of the first mountain, Kilimanjaro, when conditions were ideally standardised, the whole party was walking at the same pace and with the same objective. On Mt Kenya, activity subgroups made comparisons more difficult, but assessment over three days during which altitudes and activities were similar allowed valid comparisons to be made. On this ascent, as the table shows, low scores were achieved by both treatment groups.

The slightly lower altitude of Mt Kenya and some carry-over both of physical fitness and of acclimatisation from the ascent of Kilimanjaro may have contributed to this general improvement. It may be that taking acetazolamide on the first ascent permitted trouble-free acclimatisation and training which remained of benefit on the second ascent and that the improvement in performance experienced by those who changed to acetazolamide for the second ascent was partly due to the drug. We can only speculate on this point, however, as we did not have enough subjects to allocate a group to placebo throughout.

Subject 24 is an important exception to the general experience of benefit from acetazolamide. Clearly no general conclusion can be drawn from her case, but it would seem to indicate that not all those liable to develop severe acute mountain sickness will be helped by acetazolamide.

It seems clear that acetazolamide is a useful prophylactic for acute mountain sickness in most cases.

We wish to thank the following for help and advice: Young Explorers Trust; Lederle Laboratories, and in particular Dr P G Brock, medical director; Birmingham Medical Research Expeditionary Society; Dr Stirling, then medical officer of health of Tanzania; and British and African members of the Scout East Africa Expedition 1980.

Correspondence should be addressed to Dr A M Kerr, Ward 6A, Royal Hospital for Sick Children, Yorkhill, Glasgow.

References

1 Heath D, Williams DR. *Man at high altitude.* Edinburgh: Churchill Livingstone, 1977:118.
2 Birmingham Medical Research Expeditionary Society Mountain Sickness Group. Acetazolamide in control of acute mountain sickness. *Lancet* 1981;i:180-3.
3 Sutton JR, Houston CS, Mansell AL, *et al.* Effect of acetazolamide on hypoxaemia during sleep at high altitude. *N Engl J Med* 1979;301:1329-31.
4 Cain SM, Dunn JE. Low doses of acetazolamide to aid accommodation of men to altitude. *J Appl Physiol* 1966;21:1195-200.
5 Forwand SA, Landowne M, Follansbee JN, *et al.* Effect of acetazolamide on acute mountain sickness. *N Engl J Med* 1968;279:839-44.
6 Gray GW, Bryan AC, Frayser R, *et al.* Control of acute mountain sickness. *Aerospace Medicine* 1971;42:81-4.
7 Evans WO, Robinson SM, Horstmann DH, Jackson RE, Weiskopf RB. Amelioration of the symptoms of acute mountain sickness by staging and acetazolamide. *Aviat Space Environ Med* 1976;47:512-16.
8 Hackett PH, Rennie D, Levine HD. The incidence, importance and prophylaxis of acute mountain sickness. *Lancet* 1976;ii:1149-54.
9 Hills M, Armitage P. The two period cross-over clinical trial 1979. *British Journal of Clinical Pharmacology* 1979;8:7-20.

(*Accepted 6 July 1981*)

OF BLEEDING.

No operation of furgery is fo frequently neceffary as bleeding; it ought therefore to be very generally underftood. But though practifed by Midwives, Gardeners, Blackfmiths, &c. we have reafon to believe that very few know when it is proper. Even phyficians themfelves have been fo much the dupes of theory in this article, as to render it the fubject of ridicule. It is, however, an operation of great importance, and muft, when feafonably and properly performed, be of fingular fervice to thofe in diftrefs.

BLEEDING is proper at the beginning of all inflammatory fevers, as pleurifies, peripneumonies, &c. It is likewife proper in all topical inflammations, as thofe of the inteftines, womb, bladder, ftomach, kidnies, throat, eyes, &c. as alfo in the afthma, fciatic pains, coughs, head-achs, rheumatifms, the apoplexy, epilepfy, and bloody flux. After falls, blows, bruifes, or any violent hurt received either externally or internally, bleeding is neceffary. It is likewife neceffary for perfons who have had the misfortune to be ftrangled, drowned, fuffocated with foul air, the fumes of metal, or the like. In a word, whenever the vital motions have been fuddenly ftopt from any caufe whatever, except in fwoonings, occafioned by mere weaknefs or hyfteric affections, it is proper to open a vein. But in all diforders proceeding from a relaxation of the folids, and an impoverifhed ftate of the blood, as dropfies, cacochymies, &c. bleeding is improper.

(Buchan's *Domestic Medicine*, 1786.)