

# Clinical Trials: Task Sheet 1

## Notes & Solutions

1) Read the article referred to in §1.8, this can be accessed from the web address given there or from the link given in the course web pages. Use the facility on the BMJ web pages to find related articles both earlier and later.

Trust you have done this by now.

2) Revision of *t*-tests and non-parametric tests. And this also.

3) Using your general knowledge compare the following two theories against the Bradford-Hill Criteria:

i) *Smoking causes lung cancer*

Most of the criteria are satisfied. The weakest is whether or not there is a confounding factor that predisposes someone to smoke and that also increases the likelihood of developing lung cancer, possibly genetic. Establishing this criterion can be difficult in the absence of randomised controlled trials (out of the question with humans). The arguments against in this case are that there is evidence of passive smoking being harmful, clear evidence of links between smoking and other diseases (both other forms of cancer and non-cancer conditions such as heart disease), evidence of a link between chewing tobacco and cancers in site topically affected by tobacco juice (mouth and throat in particular).

ii) *The MMR (mumps, measles and rubella) vaccine given to young babies causes autism in later childhood.*

This theory falls on several criteria. Firstly in terms of consistency, extensive studies in other countries have failed to find evidence of such



a connection. In particular a very extensive study in Finland (I leave you to trace an account of this, try *googlescholar* and also Ben Goldacre's Bad Science web page). Secondly, specificity is not easy to establish, thirdly no plausible biological mechanism explanation has been offered.

4) For each of the proposed trials listed below, select the most appropriate study design, allocating onne design to onne trial. (Onne≡'one and only one'!)

A→b

B→a

C→d

D→c

is the best allocation subject to the constraint of onne design used onnce. Some other design might be appropriate for the situation described, e.g. C→a



5) In a recent radio programme an experiment was proposed to investigate whether common garden snails have a homing instinct and return to their 'home territory' if they are moved to some distance away.. The proposal is that you should collect a number of snails, mark them with a distinctly coloured nail varnish, and place all of them in your neighbour's garden. Your neighbour should do likewise (using a different colour) and place their snails in your garden. You and your neighbour should each observe how many snails returned to their own garden and how many stayed in their neighbour's. Full details are given at <http://downloads.bbc.co.uk/radio4/so-you-want-to-be-a-scientist/Snail-Swapping-Experiment-Instructions.pdf>

(a) What flaws does the design of this experiment have?

(b) How could the design of the experiment be improved?

(Note: this question is open-ended and there are many possible acceptable answers to both parts. Discussion is intended)

This question was set in the context of the discussion in lectures of randomized double-blind controlled trials. So the first steps are to consider what the experimental and control groups and what is the 'intervention' (i.e. the action performed by the experimenter on the test subjects which might affect the measured outcome — the intervention is performed on the experimental group but not on the control group). In this case the intervention is to move snails from their home territory and place them at some distance. The measured response is to see whether they return to their home territory. Examination of the design shows that **there is no control group**. This is a major flaw in the design of the experiment. All of the snails caught in the owner's home garden are marked and placed in the neighbour's garden. Further, all of the snails marked by the neighbour in their garden are removed to the owner's garden. If the neighbour marked their snails and then released them back in their own garden then this would be a control group (since they would not have



received the intervention). Without this control group you cannot rule out with any certainty whether snails always wander around quite a large territory covering adjacent gardens (remember the time scale is quite long – a week – between intervention and measurement of response).

A further, maybe less serious flaw, is that there is little randomization in the experiment. Presumably the snails that were captured and marked were not randomly selected from all of those in the garden but were those that were out and about and not hiding in obscure places. It is not realistic to catch all the snails in the garden and select a random sample to be exiled next door. However, a better design would be to catch say  $2N$  snails in the owner's garden, randomly select  $N$  of them to be marked with one colour and then exiled next door, the other  $N$  would be marked with a different colour and allowed to stay at home. The neighbour could reciprocate with  $2M$  snails, using two further colours. This would allow control of further potential explanatory factors such as whether snails naturally drift in one direction along the road or whether one garden is particularly attractive to snails because of the presence of young green plants in only one of the gardens and these giving off aromatic signals detectable by snails. If snails equally migrate home in both directions and none of the control groups migrate then it does suggest that the homing instinct is because of homesickness rather than seeking food or some other attraction.





## Clinical Trials: Task Sheet 2

### Notes & Solutions

1)

- i) *Fifteen individuals who attend a weightwatchers' clinic are each to be assigned at random to one of the treatments A, B, C to reduce their weights. Describe and implement a randomized scheme to make a balanced allocation of treatments to individuals.*

If using a printed table of random numbers (e.g. Neave, Table 7.1) then number people 01, . . . , 15. Take 2-digit random numbers, discard those not between 01 and 15 (fold, to make selection more efficient, if you want; then 01=21=41=61=81, etc); ignore repeats; the first 5 picked get A. Take 5 further 2-digit random numbers between 01 and 15 in the same way; ignore repeats and those that have A; these get B. The remaining 5 get C.

Taking the following random digits (Neave 7.1, row 20):

07636 04876 61063 57571 69434 14965 20911 73162`  
 Take in pairs, fold, so 01=21=41=61=81, etc. 07, 63=03, 60=20 (ignore), 48=08, 76=16(ignore), 61=01, 06. So: 07, 03, 08, 01, 06 get A. 35=15, 75=15(ignore), 71=11, 69=09, 43=03(ignore), 41=01(ignore), 49=09 (ignore) 65=05, 20 (ignore), 91=11(ignore), 17 (ignore) 31=11 (ignore) 62=02. So 15, 11, 09, 05, 02 get B. The rest get C.



If using a computer package that has a random number generator or random sample selection then there are various methods. Two are illustrated in R:

```
(a) > x<-c(1:15)
> x
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
> y<-sample(x)
> y
[1] 8 6 1 14 5 7 3 15 11 2 4 9 13 10 12
```

Then subjects 8, 6, 1, 14 and 5 are allocated to A.

```
(b)
> z<-c(rep("A",5),rep("B",5),rep("C",5))
> z
[1] "A" "A" "A" "A" "A" "B" "B" "B" "B" "B" "C" "C" "C" "C" "C"
> w<-sample(z)
> w
[1] "B" "C" "A" "A" "C" "A" "B" "C" "A" "B" "B" "A" "C" "B" "C"
```

Then the first subject is allocated to B, the second to C, etc.

ii) *Different individuals need to lose differing amounts of weight—as shown below (in pounds).*

1. 27	4. 33	7. 27	10. 24	13. 35
2. 35	5. 23	8. 34	11. 30	14. 36
3. 24	6. 26	9. 30	12. 39	15. 30

*Describe and implement a design which makes use of this extra information, and explain why this may give a more illuminating comparison of the treatments.*

Need to form blocks of similar units (here individuals); ideally, block size is the number of treatments to be compared, so here three. Hence, construct five blocks of size three. Order individuals by weight loss, and then form groups of three, giving the following blocks of individuals: (5, 3, 10), (6, 1, 7), (9, 11, 15), (4, 8, 2), (13, 14, 12); note



that '2' and '13' could be the other way round. Now assign each treatment once within each block randomly. Assign an integer to each possible order of the three treatments: 1–ABC, 2–ACB, 3–BAC, 4–BCA, 5–CAB, 6–CBA.

Taking the following random digits (Neave 7.1, row 20):

07636 04876 61063; ignoring 0, 7, 8, 9 gives 6, 3, 6, 4, 6, and so the treatments are assigned in the order: CBA BAC CBA BCA CBA. Comparisons within blocks are made over more similar individuals, thereby reducing the effect on the spread of the results of the external variable 'how much weight you need to lose'.

In R this could be achieved in a variety of ways, either with allowing different blocks to have the same order of treatments or (since only five of the six possible orderings are required) ensuring that any order is used at most once. Four are illustrated below.

```
> x<-c(1:6)
> sample(x,5)
[1] 4 1 5 6 3
> sample(x,5,replace=TRUE)
[1] 1 3 4 3 3
> y<- c("ABC", "ACB", "BAC", "BCA", "CAB", "CBA")
> sample(y,5)
[1] "BAC" "CBA" "ABC" "BCA" "CAB"
> sample(y,5,replace=T)
[1] "CBA" "ACB" "BCA" "ACB" "BCA"
```



- 2) A surgeon wishes to compare two possible surgical techniques for curing a specific heart defect, the current standard and a new experimental technique. 24 patients on the waiting list have agreed to take part in the trial; some information about them is given in the table below.

Patient	1	2	3	4	5	6	7	8	9	10	11	12
Sex	M	F	F	F	F	M	M	M	M	M	F	F
Age	64	65	46	70	68	52	54	52	75	55	50	38
Patient	13	14	15	16	17	18	19	20	21	22	23	24
Sex	M	F	F	F	M	M	M	M	M	M	F	M
Age	59	56	64	64	41	68	48	63	41	62	49	44

Devise a suitable way of allocating patients to the two treatments, and carry out the allocation.

There are lots of possible designs; randomization is vital, and balance is important (and easy to obtain). To take advantage of the extra information given, pair the patients up (because there are two treatments) as far as possible by sex and age—since both factors could affect the suitability of the treatment. The female pairs correspond to ages 38 and 46, 49 and 50, 56 and 64, 64 and 65, 68 and 70, or patient numbers 12 and 3, 23 and 11, etc. Similar pairings should be carried out for the males. Within each pair, randomize the two treatments. For example, look up digits from the beginning of Neaves table of random digits: if a pair gets a digit that is odd, assign the standard treatment to the first patient and the experimental one to the other; if they get an even digit, assign treatments the other way round.

To do this in R we need six randomly selected pairs of AB or BA:

```
> sample(c("AB", "BA"), 6, replace=T)
[1] "AB" "BA" "AB" "BA" "AB" "BA"
>
```



3) On a recent BBC Radio programme (Front Row, Friday 03/10/08, <http://www.bbc.co.uk/radio4/arts/frontrow/>) there was an interview with Bettany Hughes, a historian, (<http://www.bettanyhughes.co.uk/>) who was talking about gold (in relation to an exhibition of a gold statue of Kate Moss in the British Museum). She made the surprising statement

**"....ingesting gold can cure some forms of cancer."**

I would only regard this as true if there has been a randomized controlled clinical trial where one of the treatments was gold taken by mouth and where the measured outcome was cure of a type of cancer. The task is to find a record of such a clinical trial or else find a plausible source that might explain this historian's rash statement.

The basis of this story seems to be reports that gold nano particles have been observed to bind to receptors on certain types of cancer cells. This is a long way from saying that gold *cures cancer*. Looking on [clinicaltrials.gov](http://clinicaltrials.gov) and searching under 'gold' 'cancer' lists 80+ trials which include the two words 'gold' and 'cancer' somewhere in their protocols. Several of these use 'gold' in the phrase 'gold standard' and don't involve administering actual gold. Others seem to involve studies where gold is not claimed to be the active agent but used as a delivery vehicle for some therapeutic agent bound to colloidal gold (gold pulverised to a very fine powder). I wasn't able to find details of a couple of Phase I trials (e.g. by Mayo Clinic) but no later phases and no links to publications were given.



4) Patients are to be allocated randomly to 3 treatments. Construct a randomization list

- i) for a simple, unrestricted random allocation of 24 patients
- ii) for a restricted allocation stratified on the following factors with 4 patients available in each factor combination:

Sex: M or F      Age: <30; 30≤&<50; ≥50.

i)e.g. take 1,2,3 → A; 4,5,6 → B; 7,8,9 → C; 0 → discard. Or in R:

```
> x<-c("A","B","C")
> y<-sample(x,24,replace=TRUE)
> y
[1] "C" "B" "B" "A" "A" "A" "A" "A" "A" "A" "C" "B" "C" "C"
"A" "A" "C" "B" "A"
[20] "B" "C" "B" "A" "A"
```

- iii) Would usually take 1→ABC; 2→ACB; 3→BAC; 4→BCA; 5→CAB; 6→CBA using randomly permuted blocks of size 3. However, there are only 4 patients available at each factor combination. Possibilities are to choose 4<sup>th</sup> treatment (a) randomly or (b) selecting if one treatment is more important than the other 2 — then position that treatment randomly in the sequence (4 possible positions). Other possibilities are available.

More sophisticated in R is either:

```
lapply(rep(list(LETTERS[1:3]),4),sample)
[[1]]
[1] "B" "C" "A"

[[2]]
[1] "B" "A" "C"

[[3]]
[1] "A" "B" "C"

[[4]]
[1] "B" "C" "A"
or
```



```
matrix(apply(matrix(c("A","B","C"),3,4),2,sample),1,3*4)
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
[1,] "C"  "B"  "A"  "B"  "A"  "C"  "B"  "A"  "C"  "B"  "A"  "C"
>
```

- 5) Patients are to be randomly assigned to active and placebo treatments in the ratio 2:1. To ensure 'balance' a block size of 6 is to be used. Construct a randomisation list for a total sample size of 24.)

There 15 ( $=6!/4!2!$ ) blocks of size six of form AAAAPP. Note that a block size of 3 gives only 3 possibilities and so is unsatisfactory – too easy to crack. This can be done easily in R with `rep()` and `sample()`:

```
> sample(c(rep("A",4),rep("P",2)),6)
[1] "A" "A" "A" "P" "A" "P"
> sample(c(rep("A",4),rep("P",2)),6)
[1] "A" "A" "P" "P" "A" "A"
> sample(c(rep("A",4),rep("P",2)),6)
[1] "P" "A" "A" "A" "A" "P"
> sample(c(rep("A",4),rep("P",2)),6)
[1] "A" "A" "A" "P" "A" "P"
>
```

### More sophisticated is

```
matrix(apply(matrix(c(rep("A",4),rep("P",2)),6,4),2,sample),
1,6*4)
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
[,11] [,12] [,13] [,14]
[1,] "A"  "A"  "A"  "P"  "P"  "A"  "A"  "P"  "A"  "P"  "A"
     "A"  "A"  "P"
      [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23]
[,24]
[1,] "A"  "A"  "A"  "P"  "A"  "A"  "P"  "A"  "A"
     "P"
```



6) Patients are to be randomly assigned to active and placebo treatments in the ratio 3:2. To ensure 'balance' a block size of 5 is to be used. Construct a randomisation list for a total sample size of 30

There are 10 ( $=5!/3!2!$ ) blocks of size 5 of form AAAPP. Note that a block size of 10 of form AAAPPAAAPP would give  $10!/6!4!=210$  possibilities, perhaps too many (overkill), 10 possibilities with block size 5 is probably adequate and not easy to crack, or else take random subset of these of say 5 sets.

Either use repeatedly:

```
sample(c(rep("A",3),rep("P",2)),5)
```

```
[1] "A" "A" "P" "A" "P"
```

```
>
```

Or, more sophisticated

```
>
```

```
matrix(apply(matrix(c(rep("A",3),rep("P",2)),5,6),2,sample),
1,5*6)
```

```
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
[ ,11] [ ,12] [ ,13] [ ,14]
```

```
[1,] "A"  "A"  "A"  "P"  "P"  "P"  "A"  "A"  "P"  "A"  "A"
     "P"  "A"  "P"
```

```
      [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23]
[ ,24] [ ,25] [ ,26]
```

```
[1,] "A"  "A"  "P"  "A"  "A"  "P"  "A"  "P"  "A"
     "P"  "A"  "A"
```

```
      [,27] [,28] [,29] [,30]
```

```
[1,] "P"  "A"  "P"  "A"
```



## Clinical Trials: Task Sheet 3

### Notes & Solutions

1) A trial for the relief of pain in patients with osteoarthritis of the knee is being planned on the basis of a pilot survey which gave a 25% placebo response rate against a 45% active treatment response rate.

- i) How many patients will be needed to be recruited to a trial which in a two-sided test will detect a difference of this order of magnitude with 90% power? (Calculate this first 'by hand' and then using a computer package and compare the answers).

```
> power.prop.test(p1=0.25,p2=0.45,power=0.9,sig.level=0.05)
```

Two-sample comparison of proportions power calculation

```
n = 117.4307
p1 = 0.25
p2 = 0.45
sig.level = 0.05
power = 0.9
alternative = two.sided
```

NOTE: n is number in *each* group

```
> power.prop.test(p1=0.25,p2=0.45,power=0.9)
```

Two-sample comparison of proportions power calculation

```
n = 117.4307
p1 = 0.25
p2 = 0.45
sig.level = 0.05
power = 0.9
alternative = two.sided
```

NOTE: n is number in *each* group

So take 118 in each group.

**Note** that a significance level of 0.05 is assumed by default.



For comparison, the formula gives 115 patients in each group (230 in total), Both Minitab 13 and the program power.exe give 118 (total 236).

S-plus 6 gives the same answer to the problem which ever way you feed in the two proportions, the answer it gives is 128. This is the 'Yates continuity-corrected' value which is the default option in S-plus; changing this default in the options panel also gives 118 per group.

- ii) *With equal numbers in placebo and active groups, what active rates would be detected with power in the range 50% to 95% and group sizes 60 to 140? (Calculate for power in steps of 15% and group sizes in steps of 20).*

The program power.exe gives the following table

```
Results
-----
Two Sample test for proportions

Table of CRD calculations

      Sample size group 1
      :      60 :      80 :      100 :      120 :      140 :
-----
50 : 0.41887 : 0.39489 : 0.37872 : 0.36689 : 0.35777 :
65 : 0.45375 : 0.42488 : 0.40536 : 0.39106 : 0.38003 :
80 : 0.49491 : 0.46048 : 0.43708 : 0.41990 : 0.40661 :
95 : 0.56566 : 0.52249 : 0.49275 : 0.47073 : 0.45362 :
-----
Rows are: power

significance level = 0.05
ratio group1:group2 = 1:1
group1 proportion = .25
```

Note the obvious feature that the CRD decreases towards the top-right corner (large sample sizes, low power). This would be used to see what the chances were of detecting a range of differences for some realistic sample size and the benefits in moving to a larger sample size (at perhaps extra cost).



To do this in **R** without 20 separate calls to `power.prop.test` requires a little bit of programming but can be done quite easily. This is beyond the level of usage of **R** that most of those taking MAS361/461 will have seen or will be expected to use.

```
> group<-seq(60,140,by=20)
> power<-seq(0.50,0.95,by=0.15)
> group
[1] 60 80 100 120 140
> power
[1] 0.50 0.65 0.80 0.95
> delta<-matrix(nrow=4,ncol=5)
> for (i in 1:4) {
+ for (j in 1:5) {
+ delta[i,j]<-power.prop.test(p1=0.25,power=power[i],
+ n=group[j])$p2
+ }
+ }
> options(digits=3)
> delta
      [,1] [,2] [,3] [,4] [,5]
[1,] 0.419 0.395 0.379 0.367 0.358
[2,] 0.454 0.425 0.405 0.391 0.380
[3,] 0.495 0.460 0.437 0.420 0.407
[4,] 0.566 0.522 0.493 0.471 0.454
>
```

2) *Woollard & Cooper (1983) Clinical Trials Journal, 20, 89-97, report a clinical trial comparing Moducren and Propranolol as initial therapies in essential hypertension. These authors propose to compare the change in initial blood pressure under the two drugs.*

*Given that they can recruit only 100 patients in total to the study, calculate the approximate power of the two-sided 5% level t-test which will detect a difference in mean values of  $0.5\sigma$ , where  $\sigma$  is the common standard deviation.*

```
> power.t.test(n=50,sd=1,delta=.5)
```

```
Two-sample t test power calculation
```

```
      n = 50
  delta = 0.5
      sd = 1
sig.level = 0.05
  power = 0.6968888
alternative = two.sided
```

NOTE: n is number in \*each\* group



Note that the sample size in each group is 50 (total 100). Also note that a CRD of  $\frac{1}{2}\sigma$  means you enter the standard deviation as 1.0 and the CRD as  $\frac{1}{2}$ .

The programme power.exe gives a value for the power of 69.69%. (The formula for the approximation may give a slightly different answer).

- i) *How big a sample would be needed in each group if they required a power of 95%? (Calculate this first 'by hand' and then using a computer package and compare the answers).*

```
> power.t.test(power=0.95, sd=1, delta=.5)
```

```
Two-sample t test power calculation
```

```
      n = 104.9280
  delta = 0.5
     sd = 1
sig.level = 0.05
  power = 0.95
alternative = two.sided
NOTE: n is number in *each* group
```

Programme power.exe gives 105 in each group (210 in total).

- 3) *What evidence is there that taking fish oil helps schoolchildren concentrate?*

In summary the answer is very little evidence if any at all. A quick search on Ben Goldacre's page should lead you quickly to this article <http://www.badscience.net/2010/06/the-return-of-a-2bn-fishy-friend/#more-1675> which tells much of the story. In short, this theory has been reported widely in many newspapers (including recently The Observer, a generally well-regarded Sunday Newspaper) as proven fact. Tracing the Observer article to its source reveals that the study referred to did not involve fish oil nor was it designed to test whether it helped schoolchildren concentrate. It is salutary reading.

