St. George's Hospital Medical School

Research and Critical Skills

Term 2 Handbook

2002-3

Biomedical Sciences
Medicine

J. M. Bland

Research and Critical Skills, 2002-3

Term 2 Course Handbook and Seminar Exercises

Contents

Timetable 2

Aims and Objectives 3

Course Organisation 4

Methods of Assessment 6

Material to be Prepared in Advance of Seminars.

It is very important that you do this.

9b. Chi-squared and t methods 7

10b. Regression and Correlation 12

11b. Cohort and case-control studies **(MBBS only)** 14

12b. Analysis by intention to treat in clinical trials **(MBBS only)** 18

NOT INCLUDED HERE:

Course Notes

Research and Critical Skills

Biomedical Sciences Term 2 Timetable: 2002-3

Date	Time	Place	Topic
Tues	14.00	Lec	9a. Dr. Peacock available to help *
4 Feb	15.30	SG	9b. Chi-squared and t methods
Tues	10.15	Lec	10a. Dr. Peacock available to help *
14 Feb	11.30	SG	10b. Regression and correlation

MB BS Term 2 Timetable: 2002-3

Date	Time	Place	Topic
Fri	10.00	Lec	9a. Dr. Peacock available to help *
31 Jan	10.30/11.30	SG	9b. Chi-squared and t methods
Fri	10.00	Lec	10a. Dr. Peacock available to help *
7 Feb	10.30/11.30	SG	10b. Regression and correlation
Fri	10.00	Lec	11a. Dr. Peacock available to help *
21 Feb	10.30/11.30	SG	11b. Cohort and case control studies
Fri	10.00	Lec	12a. Dr. Peacock available to help *
7 Mar	10.30/11.30	SG	12b. Analysis by intention to treat

Seminar numbers refer to the chapters in the course book.

Lec: Everybody in G1 Lecture Theatre

SG: Seminar groups in teaching rooms. For MB BS seminar groups will be in two shifts, half at 10.30 and half at 11.30. For Bimedical Sciences there will be only one shift, at 11.30.

* Optional for those who think they need help.

St. George's Hospital Medical School

MBBS and Biomedical Sciences

First Year Course in Research and Critical Skills, 2002-3

Aims

At the end of this term, students should increase their understanding of research methods and improve their skills in critical reading of research.

Objectives for MB BS students

1. Students should be able to recognise, describe, and know the appropriate uses of:

   analysis by intention to treat in clinical trials.

   observational studies, census, sample, random sampling, effects of non-random samples, cross-sectional, case-control and cohort designs.

   small sample comparisons of means using the t Distribution, linear regression, correlation, transformations, chi-squared test for contingency tables, their assumptions and limitations, odds ratio and relative risk.

2. Students should be able to:

   read a medical paper with critical understanding of statistical issues, being aware of problems of design, analysis and interpretation.

Objectives for Biomedical Science students

1. Students should be able to recognise, describe, and know the appropriate uses of:

   small sample comparisons of means using the t Distribution, linear regression, correlation, transformations, chi-squared test for contingency tables, their assumptions and limitations.

2. Students should be able to:

   read a medical paper with critical understanding of statistical issues, being aware of problems of design, analysis and interpretation.

Research and Critical Skills, 2002-3

Course Organisation

Course Organising Team

Martin Bland (chair), Barbara Butland, Janet Peacock (all Public Health Sciences), Prof Sean Hilton (clinical representative, General Practice).

Course structure

This course continues from the Research and Critical Skills course in the Common Foundation Programme. There are four session for MBBS students and two for Biomedical Science students. This term there are no lectures, assignments or self-test sessions. Practice questions will be available on the Intranet. Teaching will be done in seminars and by your own work. You should read the course notes and one or more of the recommended books. In addition, Dr. Peacock will be available for 30 minutes before each seminar, to answer any questions you want to ask.

For each seminar, you and a small group of your colleagues will be assigned a task. This will be based on a piece of medical research and may involve questions about the medical/scientific background, the design of the study and the analysis. You can divide this between you as you wish. Each exercise starts with a list of the topics to be covered. You should read about these in the notes or books before attempting the exercise. You will then present your results in the seminar. Everyone should contribute and we expect each member to share in the presentations. You should also read in advance the questions for the other group. Remember, you will not learn anything if you do not do this work before the seminar, and if you refuse to contribute to the seminar you will be regarded as not having attended.

Seminar groups

Students are divided into seminar groups, each of which is further divided into two subgroups. Seminar group lists will be issued on a separate sheet, as will the list of tutors. Your tutors are experienced medical statisticians. If you have any problems with the course material, you should take advantage of the practical sessions to consult them. You may also consult Dr. Peacock (j.peacock@sghms.ac.uk, ext 2798, room 6.26 Hunter Wing).

Solutions to the exercises

Unfortunately, it is not practicable for us to give you written solutions for the seminar exercises. This is another reason why attending the seminars is essential.

Attendance

A record of attendance at and work in seminars and completion of assignments will be kept.

Recommended books and lecture notes

A handbook of printed notes is available for this course. You can get this from Academic Services, level 4, Hunter Wing. You will be given a copy free of charge, but if you lose this you will have to pay for a replacement. You will also find a book very useful. Be warned that there are some very bad statistics books on the market. The following basic textbooks are recommended:

Bland M, *An Introduction to Medical Statistics, 3rd. ed.* Oxford University Press, 2000.

Bland M and Peacock J, *Statistical questions in Evidence-based Medicine*. Oxford University Press, 2000. (As many practical exercises as you could possibly want.)

Campbell MJ and Machin D, *Medical Statistics: a Commonsense Approach, 3rd. ed.* Wiley, 1999.

Colton T, *Statistics in Medicine*, Little, Brown and co., 1974.

Dixon RA, Munro J, and Silcocks P, *The Evidence Based Medicine Workbook*, Butterworth/Heinemann, 1997. (A useful source of practical exercises, particularly exercises 4, 5, and 6.)

Hill AB and Hill ID, *Bradford Hill's Principles of Medical Statistics*, Edward Arnold, 1991.

Kirkwood E, *Essentials of Medical Statistics*, Blackwell, 1988.

Leaverton PE, *A review of biostatistics*, Little, Brown and co., 1991. (This programmed text is a useful aid to revision.)

Pereira-Maxwell F, *A-Z of Medical Statistics*, Arnold, 1998. (A useful dictionary-style reference.)

You should avoid like the plague *Medical Statistics Made Easy* by F B Pipkin.

Computing

You may find useful the computer aided learning program Statistics for the Terrified, written by members of the St. George's Computer Unit. Find a computer connected to the network, e.g. in the Computer Classroom, and call up Applications on the Intranet. Network Services. Click "CAL" (Computer aided learning), "General", then "Statistics for the Terrified". Most of the material covered in this program is relevant to our course.

For carrying out statistical calculations we have Arcus ProStat, EpiInfo and Clinstat available on the network, and many machines in the computer classroom have SPSS.

There is a World Wide Web page which carries practice exam questions and useful links to supporting material. From the School home page click on "Departments", "P", "Public Health Sciences", "Department Homepage", "Research and Critical Skills".

Assessment

All students will do a written exam at the end of Term 2, including either MCQ or EMI questions in Research and Critical Skills. Practice questions and answers will be available on the Intranet.

Medicine students will do a synoptic exam at the end of Term 3. Research and Critical Skills may be included in EMI questions, as a part of an integrated essay, and as a paper-reading exercise in the practical exam. It is quite likely to be in all three. The practical question will be similar to the questions in the seminars, and will use one of the papers in the Research and Critical Skills course handbook or in the bowel habit project, term 3. This will be an open book exam, so you may bring your notes with you.

Research and Critical Skills 9b

Biomedical Sciences: 4 February 2003

MB BS: 31 January 2003

Chi-squared and t methods

Read Course Notes Chapter 9 before this exercise

Objectives. At the end of this exercise students should be able to recognise,describe, and know the appropriate uses of: small sample comparisons of means using the t Distribution, transformations, chi-squared test for contingency tables, assumptions and limitations. This exercise also give practice in reading a medical paper with critical understanding of research methods issues.

We want to you to answer the questions, using the lecture notes, books, etc. You should work together as a sub-group on your question and then prepare a presentation of the answers. This will be presented to the rest of the seminar group. Have a look at the other question and think about the answer, so you will enjoy hearing the other sub-group answer it.

Sub-group 1:

The first paper concerns spontaneous pneumothorax, the presence of air or gas in the pleural space without known cause. Two treatments are compared: simple aspiration and intercostal tube drainage. Aspiration means the removal of gas by suction. Simple aspiration means insertion of a fine tube which is removed as soon as aspiration is complete. Intercostal drainage means the insertion of a tube between the ribs which is stitched in place and left for a while. There was a notable example of aspiration for pneumothorax performed on an airline passenger, with the aid of a coat hanger, a soft drink bottle and a rubber tube. Logistic regression is a method which can be used to look at the effects of several different factors simultaneously. It shows here that none of the variables measured was able to predict in which patients simple aspiration would fail. It is beyond the scope of this course. Smoking history is recorded as pack years, that is the number of 20-cigarette packs smoked per day multiplied by the number of years of smoking.

FROM THE BMJ:

The paper is by John Harvey, Robin J Prescott on behalf of British Thoracic Society Research Committee: Simple aspiration versus intercostal tube drainage for spontaneous pneumothorax in patients with normal lungs. *British Medical Journal*, volume 309, pages 1238-9, 19 November 1994.

ABOUT THIS REPORT:

(a) What kind of study is this?

(b) What shape do you think the distribution of smoking history has?

(c) In the table a series of tests of significance are given. The authors do not say what methods were used to calculate these P values. They should do so, but space is limited. What method could be used for each of the tests given?

(d) Is this study blind? Does this have any implications for the interpretation of the results?

(e) What are the conclusions and are they justified by the data?

(f) What null hypotheses are being tested in the first nine significance tests in the table (age to size of pneumothorax)? Are these useful things to test?

(g) Why would confidence intervals be a more informative approach to the analysis?

Sub-group 2:

These examples come from small scale and laboratory based studies. Such papers often contain several short investigations and little detail on the design. We will therefore concentrate on the presentation and analysis of the data.

The first example is a small clinical study of liver transplant patients, who require drugs to suppress their immune systems and prevent rejection of the new liver. Renal dysfunction is a major complication of long-term immunosuppressive therapy with calcineurin inhibitors (CNI). In this study, 28 people who had had renal dysfunction attributable to suspected CNI toxicity were randomised to either replacement of CNI with mycophenolate mofetil (study patients); or to remain on CNI immunosuppression (controls). Renal function, blood pressure, uric acid, and blood lipids were measured before and 6 months after study entry (*Lancet* 2001; 357: 587-91). The following figure was given:

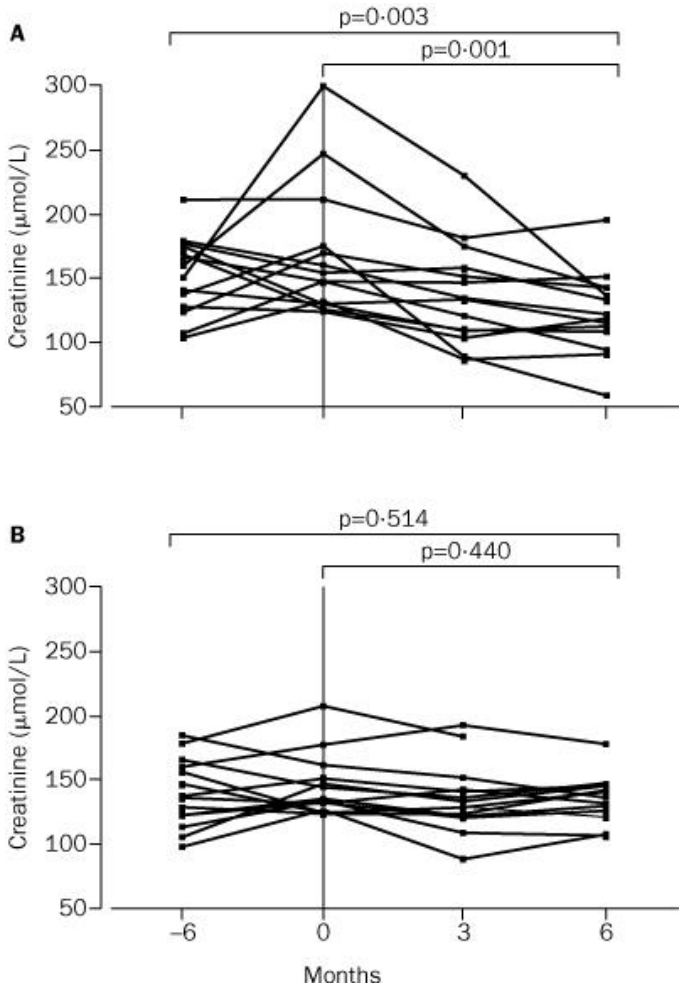

Figure: Serum creatinine concentrations in mycophenolate patients (A) and controls (B) before, at entry (0), and after study entry

The authors report: "At the end of the study, mean (SD) serum creatinine had fallen by 44·4 (48·7) µmol/L in study patients compared with 3·1 (14·3) µmol/L in controls; a mean difference of 41·3 µmol/L (95% CI 12·4-70·2)."

a) What method would be used to carry out the tests of significance shown in the figure, and why?

b) What method would be used to calculate the confidence interval, and why? What condition should the data meet for this method?

c) The standard deviations are bigger than the means. Why should we NOT conclude that change in serum creatinine has a skew distribution?

The next example (*Journal of Clinical Investigation* 2001; **108**: 1141-1150) concerned pulmonary artery smooth muscle cells (PA-SMCs) in primary pulmonary hypertension (PPH). Here the authors compared the growth of PA-SMCs from patients with PPH and from controls without PPH. The following figure appeared:

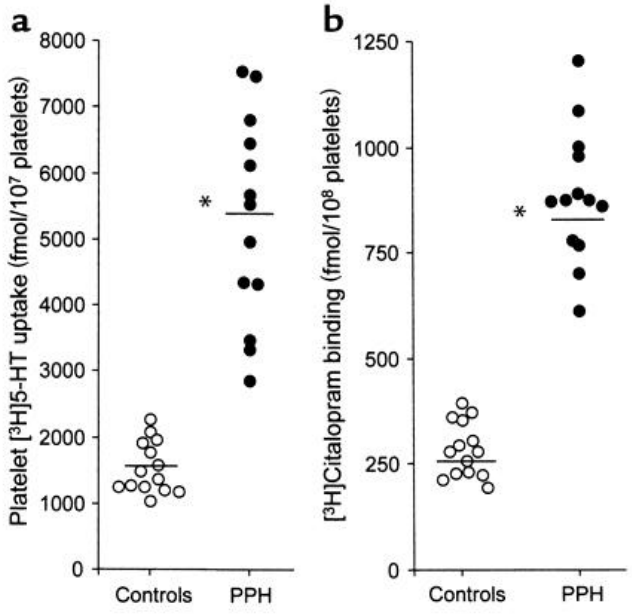

Figure. Individual platelet [3H]5-HT uptake (a) and [3H] citalopram binding (b) in normal controls (n = 14) and in patients with PPH (n = 13). The mean of each group of values is indicated by horizontal bars. * $P < 0.01$ as compared with corresponding values in controls.

d) The authors did not use a two-sample t test for these comparison, but a different test (Wilcoxon two sample test, not in our course) which does not require the same assumptions. Why did they not use a t test?

e) What could they have tried to do to make a t test acceptable?

The serotonin transporter (5-HTT) is associated with increased PA-SMC growth. The gene which controls production of 5HTT is found in two variants (alleles). The long (L) variant of the gene is associated with 5-HTT overexpression compared to the short (S) variant. People each have two copies of the gene and hence two alleles, and can be LL when both alleles are of the long type, SL when they have one long and one short, or SS when both copies are short. The following table was given:

Table 1
Genotypic distribution of serotonin transporter gene polymorphism in the promotor region in patients with PPH and in controls

	Number of subjects	Polymorphism in promoter region		
		SS	LS	LL
		n (%)	*n* (%)	*n* (%)
Normal controls	84 (100%)	16 (19%)	45 (54%)	23 (27%)
Patients with PPH	89 (100%)	7 (8%)	24 (27%)	58 (65%)

S, short variant of genotypic polymorphism, L, long variant of genotypic polymorphism

The authors commented that "The distributions of the 5-HTT genotypes in 89 patients with PPH and 84 controls were evaluated (Table 1). Patients and controls did not differ

with respect to age and body mass indices, and all subjects were white. . . . We found the genotype characterized by two long alleles in 65% of the patients with PPH and only 27% of the controls ($P < 0.001$)."

f) What statistical method should be used to calculate this P value, and why?

The third paper (*Journal of Clinical Investigation* 2001; **108**: 1031-1040) investigated an aspect of atherosclerosis, which causes most acute coronary syndromes and strokes. The pathogenesis of atherosclerosis includes recruitment of inflammatory cells to the vessel wall and activation of vascular cells. CD44 is an adhesion protein expressed on inflammatory and vascular cells. CD44 supports the adhesion of activated lymphocytes to endothelium and smooth muscle cells. To assess the potential contribution of CD44 to atherosclerosis, the authors bred CD44-null mice to atherosclerosis-prone apoE-deficient mice. They compared aortic lesions in CD44-null mice with those in CD44 heterozygote and wild-type littermates. The following figure appeared.

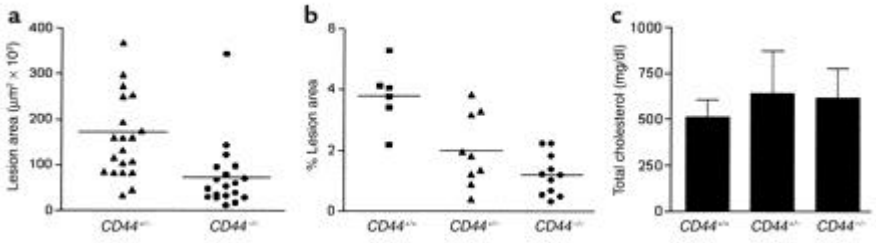

Figure. Genetic deficiency in CD44 results in marked reduction of atherosclerosis in apoE-deficient mice. (a) Atherosclerotic lesion area in CD44+/– (n = 21) and CD44–/– (n = 19) mice was quantitated by aortic root assay; P = < 0.01. (b) Atherosclerotic lesion area in the entire thoracic-abdominal aorta of CD44+/+ (n = 6), CD44+/– (n = 9), and CD44–/– (n = 10) mice quantitated by en face analysis. (c) Total cholesterol levels in the plasma of CD44+/+ (n = 18), CD44+/– (n = 20), and CD44–/– (n = 19) mice were equivalent.

g) Figure a is called a scatter plot or dot plot. What does it show?

h) Figure c is a bar chart. What might the "T" shaped vertical and horizontal lines represent.

i) Which style of figure, dot plot or bar chart, is more informative?

j) What statistical method might be used to calculate "P<0.01" for the comparison of the two groups in Figure a, and why?

*** TAKE HOME MESSAGE ***

The means of small samples can be compared using the t method, provided the assumptions of Normal distribution and uniform variance are met. When the same sample is measured twice or samples are matched, the paired t method can used provided differences are from a Normal distribution. The relationship between two categorical variables, or differences in a categorical variable between two groups, can be compared using a chi-squared test, provided the sample is large enough.

Research and Critical Skills 10b

Biomedical Sciences: 19 February 2003

MB BS: 7 February 2003

Regression and Correlation

Read Course Notes Chapter 10 before this exercise.

Objectives. At the end of this exercise students should be able to recognise, describe, and know the appropriate uses of: regression and correlation analyses, with assumptions and limitations. This exercise also give practice in reading a medical paper with critical understanding of research methods issues.

We want to you to answer the questions, using the lecture notes, books, etc. You should work together as a sub-group on your question and then prepare a presentation of the answers. This will be presented to the rest of the seminar group. Have a look at the other question and think about the answer, so you will enjoy hearing the other sub-group answer it.

Subgroup 1:

Read the following paper, "Association between raised body temperature and acute mountain sickness: cross sectional study" (*British Medical Journal*, 315, 403-4), and answer the questions.

You can ignore the references to one factor analysis of variance and Mann Whitney U test. They are equivalents of the two sample t test for more that two groups and for non-Normal data respectively.

The paper is by Marco Maggiorini, Peter Bartsch, Oswald Oelz: Association between raised body temperature and acute mountain sickness: cross sectional study. *British Medical Journal*, 315, 403-4.

QUESTIONS ABOUT THIS REPORT:

(a) Why is this a cross-sectional study?

(b) "The mean body temperature was 37.9^{o}C in climbers with cerebral oedema, compared with 36.9^{o}C in climbers with a score ≤3 (mean difference 1.0^{o}C (95% confidence interval 0.5 to 1.5))". What does this mean and what method could be used to calculate the 95% confidence interval?

(c) "The correlation coefficient between the body temperature and arterial oxygen pressure ... [was] -0.52 ($P<0.001$)". What does this mean?

(d) Figure 1 shows axillary temperature plotted against mountain sickness score, with correlation coefficients and associated P values. What condition must the data meet for the P value to be valid? Do you think the condition is met for Day 1?

(e) Do you think this study was ethical?

Sub-group 2:

The second paper describes a research exercise carried out by a GP. Epi-Info is a public domain statistical progam developed by the Centres for Disease Control, Atlanta, Georgia, for investigating outbreaks of infectious disease. It is available on the St. George's Network. Otherwise, the paper is self explanatory.

The paper is by James A Heathcote: Why do old men have big ears? (British Medical Journal, 1995; 311: 1668.

QUESTIONS ABOUT THIS REPORT:

(a) What kind of study is this?

(b) What will be the effects of using patients attending their general practice?

(c) Will the dropping of patients due to the seriousness of their presenting problem or the late running of the surgery have any effect?

(d) Is this study blind? Does this have any implications for the interpretation of the results?

(e) Is the distribution of ear size skew or symmetrical, and why?

(f) What is a regression equation? What does the one in the paper tell us? Can we conclude that the mean ear size at birth is 55.9 mm?

(g) What assumptions about the data are required for regression analysis and do you think they satisfied here?

(h) What are the conclusions and are they justified by the data?

(i) What further investigations could be done?

*** TAKE HOME MESSAGE ***

The relationship between two continuous variables can be estimated using regression, provided deviations from the line follow a Normal with uniform variance. It is dangerous to extrapolate beyond the data. The strength of the relationship can be measured by a correlation coefficient.

Research and Critical Skills 11b

MB BS only: 21 February 2003

Cohort and case-control studies, odds ratios and relative risks

Read Course Notes Chapters 11 and 12 before this exercise.

Objectives. At the end of this exercise students should be able to recognise,describe, and know the appropriate uses of: cohort study design, case-control study design, cross-over design, relative risk, odds ratio. This exercise also give practice in reading a medical paper with critical understanding of research methods issues.

We want to you to answer the questions, using the lecture notes, books, etc. You should work together as a sub-group on your question and then prepare a presentation of the answers. This will be presented to the rest of the seminar group. Have a look at the other question and think about the answer, so you will enjoy hearing the other sub-group answer it.

In this exercise each group will look at two studies. One is a case-control study and the other is not.

Sub-group 1:

In a study of functional tests for early osteoarthritis, a random sample of 2000 people aged 35-54 years living in a defined geographical area of Sweden were sent a questionnaire about knee pain. Of the 1853 (93%) who responded, 279 reported long-standing knee pain (sic) and were offered an examination. 264 (95%) accepted, of whom 60 were diagnosed as having osteoarthritis of the knee and 204 were not. These 264 patients underwent a series of functional tests, including balance tests (standing on one leg), one-leg-rising test (rising from a sitting position using one leg only, number done in fixed time), and walking (time to walk 300m).

The following table was included in the results:

Odds ratios between those with and those without radiographic tibiofemoral OA for each test (95% confidence intervals are included for each ratio)

Test	*Odds ratio*	*95% Confidence intervals*
Balance I	1.07	-0.02; 2.16
Balance III	1.14	0.24; 2.04
OLR >15	1.43	0.49; 2.37
OLR >10	1.62	0.66; 2.58
Walking < 2 min 50 sec	0.78	-0.17; 2.39
Walking > 2 min 30 sec	1.13	-0.08; 2.34

OLR = one-leg-rising test.

(*Physiotherapy Research International* 1998; **3**: 153-163)

(a) What is meant by "a random sample"? What populations do the sample of 2000 sent the questionnaire and the sample of 264 knee pain sufferers represent?

(b) Which of the following terms do you think best describes the study: case report, case series, case-control study, clinical trial, cohort study, cross-sectional study. Why is this method appropriate here?

(c) What is meant by "odds" and by "odds ratio"? What does an odds ratio of 1.07 tell us about the Balance I test as a predictor of osteoarthritis?

(d) What is wrong with the confidence intervals?

Forty five alcoholic patients and 23 non-alcoholic research or laboratory staff were studied. A two allele polymorphism was identified and subjects were classified into three genotypes, AA, AB and BB. The following table was produced:

	Number of patients with genotype		
	BB	AB	AA
Alcoholic patients	19	19	7
Non-alcoholic subjects	2	3	18

χ^2=25.8, P<0.001 (df=2)

(*BMJ* 1993;307:1388-90)

(e) What kind of study is this? Why is this design particularly suitable here?

(f) To what problems might the use of data obtained from research or laboratory staff lead?

We can calculate odds ratios for alcoholism and genotype, comparing BB to AA, and for alcoholism and genotype comparing AB to AA. These are:

	BB	AA	Total
Alcoholic	19	7	26
Non-alcoholic	2	18	20

Odds ratio = (19×18)/(7×2) = 24.4

95% confidence interval for odds ratio: 4.5 to 133.5

	AB	AA	Total
Alcoholic	19	7	26
Non-alcoholic	3	18	21

Odds ratio = (19×18)/(7×3) = 16.3

95% confidence interval for odds ratio: 3.6 to 72.9.

(g) What do these odds ratios and their confidence intervals tell us?

Sub-group 2:

In a study of alcohol consumption and mortality, men aged 35-64 were screened in 1970-3 at 27 workplaces in the west of Scotland. 5766 men who answered questions on their usual weekly alcohol consumption had mortality from all causes, coronary heart disease, stroke, and alcohol related causes recorded over 21 years of follow up, The risk for all cause mortality was similar for non-drinkers and men drinking up to 14 units a week. Mortality risk then showed a graded association with alcohol consumption (relative risk compared with non-drinkers 1.34 (95% confidence interval 1.14 to 1.58) for 15-21 units a week, 1.49 (1.27 to 1.75) for 22-34 units, 1.74 (1.47 to 2.06) for 35 or more units). Adjustment for other risk factors attenuated the increased relative risks, but they remained significantly above 1 for men drinking 22 or more units a week. There was no strong relation between alcohol consumption and mortality from coronary heart disease after adjustment. A strong positive relation was seen between alcohol consumption and risk of mortality from stroke, with men drinking 35 or more units having double the risk of non-drinkers, even after adjustment. The authors concluded that the overall association between alcohol consumption and mortality is unfavourable for men drinking over 22 units a week, and there is no clear evidence of any protective effect for men drinking less than this. (*BMJ* 1999; 318:1725-1729)

(a) What kind of study is this?

(b) What are the advantages of this design to study alcohol consumption and mortality?

(c) What is a "relative risk" and what is implied by a relative risk of 1.34?

(d) What method would be appropriate to calculate relative risk in this study?

Infants who died from Sudden Infant Death Syndrome (SIDS or cot death) were compared to a group of live infants, matched for age and birthweight. The temperature in the baby's bedroom and the amount of thermal insulation (clothes and bedding) were measured, to give an estimate of the excess thermal insulation. The dead children had had more excess thermal insulation (mean 2.3 togs, standard deviation 3.4 togs) than the live children (mean 0.6 togs, standard deviation 2.3 togs), a significant difference, P=0.009. Infants who died were also more likely than live infants to have been laid to sleep in the prone position (odds ratio 4.58, 95% confidence interval 1.48 to 14.11).

(*BMJ* 1992;304:277-282)

(e) What kind of study is this and what are the advantages and disadvantages of this design to study cot deaths?

(f) What is meant by "odds ratio" and what does an odds ratio of 4.58 tell us?

(g) Excess thermal insulation has the standard deviation (3.4 togs) greater than the mean (2.3 togs). Why can we NOT conclude that this distribution is skew?

*** TAKE HOME MESSAGE ***

Cohort studies start with the measurement of the risk factor on a group of people, who are then followed to see whether the desease develops. They suffer from fewer problems of interpretation than case-control studies. Case-control studies start with a group of people with the disease, who are then compared to another group without the

disease, the controls. They are relatively quick and cheap, but there may be problems due to the selection of cases and controls, and due to recall of information from the past. Cross-sectional studies measure disease and risk factor at the same time. They are suitable for common, chronic diseases.Risk is the probability that the disease or other event will happen. Relative risk is the ratio of two risks, how many times more likely the disease is for people in one group than it is in another.

Odds is a different way of recording probability. The odds ratio is a different way of comparing probability or proportions between two groups. It has a special interpretation case-control studies, though it can be used in many other situations.

Case-control studies start with the disease, so we can't estimate risks. We can estimate relative risk from the odds ratio.

Research and Critical Skills 12b

MB BS only: 7 March 2003

Analysis by intention to treat

Read Course Notes Chapter 13 before this exercise.

Objectives. At the end of this exercise students should be able to recognize,describe, and know the appropriate uses of: analysis by intention to treat. It provides revision in the concepts of significance tests and P values, standard errors, and large sample comparisons of two proportionsand two means.

We want you to answer the questions, using the lecture notes, books, etc. You should work together as a sub-group on your question and then prepare a presentation of the answers. This will be presented to the rest of the seminar group. Have a look at the other question and think about the answer, so you will enjoy hearing the other sub-group answer it.

The Know Your Midwife (KYM), carried out at St George's, was very influential in changing the way maternity care is delivered. The idea was that one midwife would do all the ante-natal checks, deliver the baby, and give the post-natal care. At that time, a women might see a different midwife at every stage. We wanted to compare the KYM scheme with the then standard care. The NHS would only agree to fund the KYM service on an experimental basis if an evaluation was done.

The standard way to do this would be to invite women to participate in the trial, explaining the alternative treatments, so obtaining informed consent to randomization. We would then randomize women to KYM or standard care.

The originator and principal researcher, Caroline Flint, was a great advocate of and enthusiast for continuity of care. She thought that few women who knew about the KYM scheme would accept anything else. If this was so, we would be asking people into a trial where they would very much want one treatment and be disappointed if they didn't get it. This would be bad for them and might affect the results. This problem was solved by the use of what is now called a randomized consent design. Women were randomized into two groups, "offer KYM" and "standard care", without their knowledge. The Know Your Midwife scheme was then offered to one group. They could refuse this and opt for standard care. Those who opted for KYM became the Acceptors group, those who opted for standard care became the Refusers group. The other group did not have the option and received standard care. They became the Control group. All the women were later asked to provide interview data, for which they gave consent in the usual way. They were asked to participate in a study of maternity services, but not told that they had been randomized.

Thus there were three groups of women:

Acceptors: women who were offered and received KYM.
Refusers: women who were offered but refused KYM, received standard care.
Controls: women who were offered and received standard care.

The tables show the numbers of babies delivered by Caesarean section and the mean fall in haemoglobin from before to after delivery. The latter provides an indication of the loss of blood during the delivery. For some women one of these measurements was missing, so there are fewer women in this table.

Incidence of Caesarean delivery in the KYM trial:

Mode of delivery	Refusers	Acceptors	Controls
Caesarean	7 (16%)	30 (7%)	35 (7%)
Vaginal	36 (84%)	406 (93%)	438 (93%)
Total	43 (100%)	436 (100%)	473 (100%)

Mean fall in haemoglobin in the KYM trial:

	Refusers	Acceptors	Controls
Mean	0.53	0.31	0.36
Standard deviation	1.3	1.3	1.3
Number of women	30	357	393

Subgroup 1

(a) Is the randomized consent design ethical in this trial?

(b) We could use the large sample method for the comparison of two proportions to compare the proportion of Caesareans between the Refusers and Acceptors. This gives P=0.03. Is this significant? What does this test tell us?

(c) Could a difference between these two groups possibly be due to the difference in treatment?

(d) Could a difference between these two groups possibly be due to the selection into refusers and acceptors, i.e. could refusers differ from acceptors in ways other than the treatment?

(e) At the time when the women were allocated into the three groups, i.e. before treatment began, were any two groups comparable? In other words, if they had all received the same treatment, would we expect the same outcome in each group?

(f) The authors combined the Acceptors and Refusers to give a group of all those allocated to KYM, which they then compared to the Controls. Why did they do this?

(g) The combined KYM groups comprise 479 women of whom 37 (8%) had Caesareans. Comparing this combined group with the controls we get P=0.8. Is this difference significant? What can we conclude?

Subgroup 2

(a) We could use the large sample method for the comparison of two means to compare the mean change in haemoglobin between the Refusers and Acceptors. This gives P=0.4. Is this significant? What does this test tell us?

(b) Could a difference between these two groups possibly be due to the difference in treatment?

(c) Could a difference between these two groups possibly be due to the selection into refusers and acceptors, i.e. could refusers differ from acceptors in ways other than the treatment?

(d) At the time when the women were allocated into the three groups, i.e. before treatment began, were any two groups comparable? In other words, if they had all received the same treatment, would we expect the same outcome in each group?

(e) The authors combined the Acceptors and Refusers to give a group of all those allocated to KYM, which they then compared to the Controls. Why did they do this?

(f) The combined group allocated to KYM has mean 0.33 and standard deviation 1.30. If we test the difference between the combined group and the controls, we get P=0.7. Is the difference significant? What can we conclude?

(g) Change in haemoglobin is available for only 82% of women. How might this affect the results?

*** TAKE HOME MESSAGE ***

When people drop out of a trial after randomization, or do not get the treatment to which they were allocated for any reason, we must analyse the trial by intention to treat, keeping subjects in the group to which they were allocated originally. This analysis can be made more difficult by missing data, particularly when the proportions with missing data differ between the treatment groups.

We can assess the strength of the evidence for a treatment effect by a significance test. This gives a P value, which is the probability of getting a difference as large as that observed if the null hypothesis were true, i.e. if there were no difference between treatments in the whole population. A small P value means there is strong evidence that a difference exists in the population as a whole. A large P value means there is only weak evidence that a difference exists in the population. It does NOT mean that there is no difference.

References

Flint, C. and Poulengeris, P. (1986) The `Know Your Midwife' Report, Caroline Flint, London

Flint C, Poulengeris P, Grant A. The `Know Your Midwife' scheme -- a randomised trial of continuity of care by a team of midwives. Midwifery 1989; 5: 11-16.