

**St. George's Hospital Medical School**



# **Research and Critical Skills**

**Term 3 Handbook**

**2002-3**

**MB BS**

**J. M. Bland**

# **Research and Critical Skills, 2002-3**

## **Course Handbook and Seminar Exercises**

### **Contents**

Timetable	2
Aims and Objectives	3
Methods of Assessment	4

### **Material to be Prepared in Advance of Seminars.**

#### **It is very important that you do this.**

13b. Multiple significance tests	5
14b. Combining data from many studies	8
15b. Examination preparation	13

#### **NOT INCLUDED HERE:**

**Course notes**

## **Research and Critical Skills**

### **MB BS**

#### **Timetable: 2002-3**

<b>Date</b>	<b>Time</b>	<b>Topic</b>
Fri 26 Apr	3.00/4.00	13b. Multiple significance tests
Fri 10 May	3.00/4.00	14b. Combining data from many studies
Fri 7 June	3.00/4.00	15b. Examination preparation

All sessions are in small groups. There are no lecture theatre sessions this term. Seminar groups will be in two shifts, 4-9 at 3.00 and 10-15 at 4.00.

## St. George's Hospital Medical School

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

#### Aims

At the end of this course, students should:

1. understand the need for critical appraisal of the medical literature.
2. have a strategy for critical appraisal of a paper.
3. have a basic knowledge of statistical methods and principles used in the medical literature and of their appropriate use.

#### Objectives (specific objectives for Term 3 in bold)

1. Students should be able to recognise, describe, and know the appropriate uses of:  
clinical trial, volunteer and historical bias, randomization, **cross-over design**, placebo effect, assessment bias, double-blind technique, analysis by intention to treat.  
observational studies, census, sample, random sampling, effects of non-random samples, cross-sectional, case-control and cohort designs.  
types of data, frequency distribution, histogram, mean, median, centile, range, variance, standard deviation, 95% reference interval or normal range.  
rates and proportions, missing denominators, bar charts, scatter diagrams, line graphs, problems of missing baselines, broken bars, etc., log scales.  
probability, its simple properties, the Binomial and Normal Distributions, their role in statistics.  
standard error, confidence intervals, significance tests, **multiple significance tests**, standard errors and related large sample confidence intervals and significance tests for a sample mean, for a proportion, for differences between means and proportions.  
small sample comparisons of means using the t Distribution, sign test, linear regression, correlation, transformations, chi-squared test for contingency tables, assumptions and limitations, odds ratio and relative risk.  
**systematic review and meta-analysis.**
2. Students should be able to:  
read a medical research paper with critical understanding of statistical issues, being aware of problems of design, analysis and interpretation.

### **Assessment for Term 3**

There will be a written exam at the end of Term 3, including either MCQ or EMI questions in Research and Critical Skills. This will cover the learning objectives for Term 3 and may also include material from earlier in the course which is relevant to the Term 3 topics.

There will be 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. There is no specific essay practice in this course (it is provided in PBL), but a specimen question is given here. 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 handbooks or in the Term 3 Bowel Habit Project. This will be an open book question, so you may take your course notes with you. A specimen question is included here. Extra practice questions will be put on the course web page. This is found under Departments, Public Health Sciences.

## Research and Critical Skills 13b

26th April 2003

### Multiple significance tests

**Read Course Notes Chapter 15 before this exercise.**

*Objectives. To consider problems of multiple significance tests. To see that an understanding of statistics helps us to read the medical literature, and to see that this must be done critically. To revise some basic ideas of medical statistics.*

In this exercise we shall look critically at two papers. Both contain an example of multiple testing. Be sure to read both papers in advance.

#### Subgroup 1:

This study caused a great storm. It appeared to show that eating genetically modified food may be harmful. Its results were revealed in a television programme before it had been submitted for publication in a peer-reviewed journal, Dr Pusztai lost his job, a special commission of the Royal Society studied it and concluded that it was methodologically flawed. (Of course, almost all medical research is methodologically flawed. The reader needs to be able to spot the flaws and judge how they limit the conclusions which can be drawn.) Then the *Lancet* published it.

This paper is a tough read, so I'll summarize. Rats were fed on three kinds of potato, with or without a toxic lectin derived from snowdrops (*Galanthus nivalis*):

- unmodified, (plain and simple, the parent potatoes as can be bought in Sainsbury's)
- parent potatoes to which extra lectin had been added (+GNA)
- potatoes which had been genetically transformed to produce their own lectin (GNA-GM).

Each kind of potato was given boiled or raw, making six diets altogether. Each diet was fed to six rats. The rats were killed and five sections of the gut were examined. The mucosal thickness was measured in the stomach and the crypt length was measured in each of the other sections. The following simple comparisons were done:

- boiled vs. raw for each of the three potato/lectin combinations separately
- parent vs. parent + lectin for boiled and for raw potatoes separately
- parent + lectin Vs modified for boiled and for raw separately
- parent vs. modified for boiled and for raw separately.

Then some more complicated comparisons were done:

- lectin vs. no lectin (effect of GNA) averaged over cooking types,
- boiled vs. raw (effect of cooking) averaged over potato types
- modified vs. parent plus lectin (effect of transformation) averaged over cooking types.

You can assume that the statistical methods are correct. I have simplified this paper by cutting out some technical material and some analysis which is not mentioned in the discussion. I have also put in some headings.

**The paper is by Stanley W B Ewen, Arpad Pusztai: Effect of diets containing genetically modified potatoes expressing *Galanthus nivalis* lectin on rat small intestine. *Lancet* 1999; 354: 1353 - 1354.**

### **QUESTIONS ABOUT THIS REPORT:**

- (a) What is the main research question in this study?
- (b) How many tests of significance are there in Table 1? Why should this make us cautious?
- (c) How many tests are there comparing boiled with raw potatoes? How many of these are significant?
- (d) Do you think there is good evidence that boiling the potatoes affects the gut of rats? Does it matter to main research question?
- (e) Which tests in Table 1 look for the effect of lectin?
- (f) Do you think there is good evidence that adding lectin to a potato diet affects the gut of rats? Does it matter to main research question?
- (g) Which tests in Table 1 look for the effect of the genetic transformation process, apart from the effect of the lectin?
- (h) Do you think there is good evidence that the genetic transformation process affects the gut of rats? (NB: the P value for colon, raw, Parent+GNA vs. GNA-GM, given as  $<0.01$ , is in fact 0.003, calculated from the data quoted in the table.)
- (i) What would you conclude about the effects of genetic modification from the evidence here?

### **Subgroup 2:**

This paper concerns an attempt to prevent bone loss in older women. After the menopause, women's bones tend to become less dense and more brittle (osteoporosis), leading to an increased risk of fractures.

One possible way to combat the demineralisation is to give a diet high in calcium or calcium supplements. Another is to encourage exercise.

Read the paper and then answer the questions.

**The paper is by Marion E T McMurdo, Patricia A Mole, Colin R Paterson: Controlled trial of weight bearing exercise in older women in relation to bone density and falls. *British Medical Journal*, 1997; 314: 569.**

### **Questions about this report**

- (a) What kind of study is this?
- (b) What confidence intervals would be more useful than the ones in Table 1?
- (c) "We found no significant differences between the groups at baseline." Why were these tests of significance unnecessary?
- (d) The authors state that "The difference between the groups in the number of women falling during the whole two year period was not significant ( $P=0.158$ ), but between 12 months and 18 months into the study the difference was significant ( $P=0.011$ )."  
What method could be used to find " $P=0.158$ "? What analysis would be more informative?
- (e) Does " $P=0.011$ " add anything of value?

(f) What can we conclude from this study?

**\*\*\* TAKE HOME MESSAGE \*\*\***

Significance tests must be interpreted with caution. When many tests are done and one significant one is reported, there is a danger that this may be the 1 in 20 which are significant by chance. The medical literature is one of the main resources we have for deciding on treatments, diagnostic methods and preventive measures. Statistics is a fundamental tool of medical research and papers often depend upon it. Research is difficult and the people who carry it out are not infallible. The reader must be able to use judgement about the results of studies, even when published in prestigious journals.

## Research and Critical Skills 14b

10 May 2003

### Combining data from many studies

**Read Course Notes Chapter 16 and revise Course Notes Section 2.5 from the CFP Research and Critical Skills handbook before this exercise.**

*Objectives.* After this exercise and associated reading, you should understand two approaches to evaluating all the available evidence on an issue: systematic review and meta-analysis, the need for completeness of study identification and data, heterogeneity, presentation of meta-analysis. You should also know the appropriate use of cross-over designs.

In this exercise we shall look critically at two reviews of the available evidence. One is a systematic review of the evidence about the complementary therapy acupressure, the other is a meta-analysis of the effectiveness of antibiotics in the treatment of acute cough in adults. Because such papers tend to be long, I have shortened them considerably. I have also omitted the references.

#### **Subgroup 1:**

This paper is taken from the journal *Obstetrics & Gynecology* (Vol. 91, No. 1, January 1998, pages 149-155). The original paper covers all complementary therapies. I have edited it to concentrate on only one: acupressure (acupuncture without needles) on the wrist. (Fans of Xena Warrior Princess may recall that she cured her friend's seasickness by this method.)

Read the paper and answer the questions.

#### **ALTERNATIVE THERAPIES FOR NAUSEA AND VOMITING OF PREGNANCY**

*Patricia Aikins Murphy, CNM, DRPH*

*Objective.* To review available evidence about the effectiveness of alternative therapies for nausea and vomiting of pregnancy.

*Data Sources:* MEDLINE and 13 additional US and international data bases were searched in 1996-1997 for papers that described use of alternative medicine in the treatment of pregnancy and pregnancy complications, specifically those addressing nausea, vomiting, and hyperemesis. Bibliographies of retrieved papers were reviewed to identify additional sources.

*Methods of Study Selection:* All relevant English language clinical research papers were reviewed. Randomized clinical trials addressing specifically the use of nonpharmaceutical and nondietary interventions were chosen for detailed review.

*Tabulation, Integration, and Results:* Ten randomized trials studying the effects of acupressure, ginger, and pyridoxine on nausea and vomiting of pregnancy were reviewed. Evidence of beneficial effects was found for these three interventions, although the data on acupressure are equivocal. Insufficient evidence was found for the benefits of hypnosis. Other interventions have not been studied.

*Conclusion:* There is a dearth of research to support or to refute the efficacy of a number of common remedies for nausea and vomiting of pregnancy. The best-studied alternative remedy is acupressure, which may afford relief to many women; ginger and vitamin B. also may be beneficial.

*From the Department of Obstetrics and Gynecology, and The Center for Complementary and Alternative Medicine Research in Women's Health, Columbia University College of Physicians and Surgeons, New, York, New, York.*

*This project was supported through a National Institutes of Health, Office of Alternative Medicine grant U24-HD33199.*

Nausea and vomiting of early pregnancy is a common complaint, affecting 50% or more of women in Western societies. The condition is generally self-limited, with symptoms most common and troublesome in early pregnancy and abating by the end of the first trimester. Concerns about the potential teratogenic effects of drugs taken during the critical embryogenic period, especially in the wake of the thalidomide tragedy of the 1960s, limit the use of pharmacologic treatments for this condition. As a result, many women try alternative therapies to treat nausea and vomiting. An informal review of a number of pregnancy self-help books, magazines, and lay alternative therapy publications indicates that therapies range from vitamins to herbals to homeopathic drugs to acupressure or acupuncture.

This review summarizes research on a variety of complementary and alternative remedies for the treatment of nausea and vomiting of pregnancy. [Only acupressure is given here.] Published studies are too few or have used an assortment of exposures and outcome measures, making a statistical meta-analysis difficult, if not inappropriate.

## *Methodology*

A literature search was conducted in 1996, as part of a broad strategy to identify relevant publications relating to the use of complementary or alternative medicine in women's health. References for this review were culled from the overall search results. Databases searched included, in addition to MEDLINE, Acubase (acupuncture), Agricola (National Agricultural Library, USA), Biosis (USA), CATS/AMED (Current Awareness Topics/Alternative and Allied Medicine Database, UK), Cancerline, Cinahl, CISCOM (Centralised Information Service for Complementary Medicine, UK), Embase (coverage of pharmacologic and biomedical research, the Netherlands), and the General Science, Psych Abstracts, Psych Info, and Social Science Citations Indexes.

To identify any additional references not found in the automated search, citations and bibliographies of all retrieved papers were reviewed as well. The search was not limited to English language papers; however, only English language papers or abstracts were read.

Papers were included in this review if they discussed 1) nausea and/or vomiting of early pregnancy as the treatment condition and 2) an intervention or therapy other than Western biomedical pharmacology or standard advice about diet or lifestyle changes. All research studies were reviewed; studies were classified as clinical trials (comparing clinical outcomes of women given the alternative intervention with a control group), observational studies, or case reports or series. Decisions were made *a priori* to exclude papers that represented pooled data from other studies, if the original studies were published in English and retrieved in the literature search.

## *General Considerations in Evaluating Studies*

Issues to consider when evaluating studies of alternative therapies include 1) the highly individualized nature of treatment in many alternative frameworks, 2) the difficulty of creating appropriate placebos for some interventions, and 3) the reliability and validity of exposure and outcome measures. In addition, because nausea and vomiting of pregnancy is a self-limited condition, some women will get better during the period of therapy with or without the intervention. Thus, the effects of time and gestational age must be considered when studying interventions. Many women try multiple remedies, and it often is difficult to distinguish their effects; studies should demonstrate some effort to control for other therapies. In randomized trials, distribution of other interventions might be assumed to be equal across groups, but in other studies, their effects should be controlled. Finally, consideration should be given to the potential for adverse effects on fetal development, given that most of these interventions are used during a critical period of embryogenesis. Additional remedy-specific methodologic considerations will be discussed in the individual sections below.

**Table 2. Randomized Clinical Trials of Acupressure-**

Study	Number of subjects	Intervention	Results
Dundee <i>et al</i> , 1988 <sup>3</sup>	350	Manual acupressure over the P6 or a dummy point, compared with control group with no intervention. Intervention of 4 days' duration.	“Severe” or “troublesome” morning sickness noted in 24% using P6 acupressure, 37% using dummy point acupressure, and 56% using no intervention ( $p < .005$ ).
Hyde, 1989 <sup>4</sup>	16	Acupressure wristbands vs. no intervention. 5 days intervention, followed by crossover to other group for 5 days.	Relief of morning sickness in 75% using wristbands ( $P < .025$ ); reduction of anxiety, depression, and behavioral dysfunction, as measured by standard psychometric tools ( $P < .05$ ).
DeAloysio <i>et al</i> , 1992 <sup>5</sup>	60	Unilateral (right or left wrist) or bilateral acupressure wristbands* vs. placebo wristbands (no pressure exerted on P6 point). Each version of intervention given for 3 days in a crossover design.	Reduction or elimination of symptoms in 65-69% while using acupressure vs. 29-31% while using placebo ( $P < .05$ ).
Bayreuther <i>et al</i> , 1994 <sup>6</sup>	16	Acupressure wristbands* vs. placebo wristbands (applied over a dummy point). Seven-day intervention followed by crossover to the alternative intervention for 7 days.	Nausea score (using a visual acuity scale) was lower in the group using acupressure compared with the placebo ( $P = .019$ ). No effect on vomiting.
Belluomini <i>et al</i> , 1994 <sup>7</sup>	60	Manual acupressure on the P6 point vs. pressure on a dummy point for 3 days.	Nausea decreased significantly in treatment group ( $P = .0021$ ). No difference in severity or frequency of vomiting.
O'Brien <i>et al</i> , 1996 <sup>8</sup>	161	Acupressure wristbands* vs. wristbands over a dummy point vs. a control group with no intervention. Seven-day intervention.	No differences across groups in nausea or vomiting.
Evans <i>et al</i> , 1993 <sup>9</sup>	23	Continuous electric current stimulation at the P6 point vs. no stimulation. Forty-eight-hour intervention followed by crossover to alternative intervention.	Improvement in symptoms of nausea and vomiting in 87% of experimental group vs. 43% of controls ( $p < .05$ ).

P6 = pericardium 6.

\* Sea Bands (Sea Band International, Greensboro, NC).

### Acupressure and Related Modalities

There have been several clinical trials<sup>3-9</sup> of acupressure and related modalities in the treatment of nausea and vomiting of pregnancy. All involved stimulation of or pressure on an acupuncture point known as pericardium 6 (P6) or the Neiguan point, on the volar surface of the forearm approximately three fingerbreadths above the wrist. A review<sup>10</sup> of studies of this intervention has demonstrated its beneficial effect in the treatment of postoperative emesis, chemotherapy-associated emesis, and motion sickness.

One notes that it probably is impossible to perform a true double-blind trial compared with no intervention. Because of the nature of acupressure or acupuncture, subjects will be aware that the

intervention is being applied, and any reduction in symptoms could be due to the placebo effect. Sham acupressure (applying pressure at a point other than the one thought to be important) has been used in some trials. However, the correct point can be identified easily in any number of self-help books. In addition, even sham acupressure may elicit a therapeutic response.<sup>11</sup> Finally, these studies<sup>3-9</sup> represent a variety of intervention and placebo approaches, differing treatment lengths, varying outcomes measures, and an assortment of outcome categories. Therefore, statistical meta-analysis was deemed inappropriate.

Other studies were identified,<sup>12-14</sup> but they did not meet the criteria for randomized controlled clinical trials. Those reporting benefits were essentially single-group observational studies<sup>12-14</sup> with no controls and examined both acupressure and acupuncture. A randomized trial of acupuncture for the treatment of nausea and vomiting in early pregnancy is under way in Australia (personal communication, Caroline Smith, 1997).

## *Discussion*

The best-studied remedy appears to be acupressure over the Neiguan point. A systematic review of this literature,<sup>30</sup> which excluded studies of poor methodologic quality and those whose data could not be combined, found the effects of acupressure to be comparable to those of anti-emetic medications but cautioned that the evidence is equivocal. Indeed, clinical trials on the use of acupressure or related modalities have not produced consistent findings, although before the publication of the study by O'Brien and colleagues,<sup>8</sup> results of published trials<sup>3-7,9</sup> were largely positive. It is clear that acupressure of the Neiguan point benefits many women. One could infer that in studies that did not provide an intervention to the control group or arm of the trial, positive findings might be due only to the placebo effect, because of the difficulty of creating a true double-blind trial. However, there may be value in mobilizing such a mind-body interaction if the intervention is simple and inexpensive, if it results in reduction of symptoms, and if it is not associated with any risk. Properly applied pressure would seem to produce little risk of adverse effect. There are acupressure and acupuncture points that are contraindicated in pregnancy because of their potential to produce uterine contractions, but these are not near the Neiguan point. Such an intervention also would be inexpensive to implement and readily available without a medical visit. No evaluable trials of the related modality, acupuncture, were found. The latter intervention would be more costly, in terms of time and travel for treatment visits and provider fees.

In summary, women seeking alternative, nonpharmacologic therapies for nausea and vomiting of early pregnancy have few evidence-based guidelines to assist them. They may be advised to try acupressure over the Neiguan point; this can be achieved through the use of commercially available wristbands or by applying manual pressure to the appropriate spot on the volar surface of the wrist. [The paper reported studies which suggested that ginger and vitamin B<sub>6</sub> may also be efficacious.] Women should be cautioned that apart from these few studies, there is little evidence to support or refute the benefits or risks of other remedies.

Copyright 1998 by The American College of Obstetricians and Gynecologists.

Published by Elsevier Science Inc.

References omitted.

## **QUESTIONS ABOUT THIS REPORT:**

- (a) What kind of study is this?
- (b) Why does the author think that statistical meta-analysis would be inappropriate? Do you agree?
- (c) Which of the trials in the Table are crossover trials?
- (d) What are the arguments for using crossover trials to study this topic?
- (e) What are the arguments against using crossover trials to study this topic?

- (f) What problems are caused by the results being reported in terms of P values?
- (g) What are the author's conclusions concerning acupressure and are they justified by the data?

### **Subgroup 2:**

This report is from the *BMJ*. I have edited it to shorten it a bit. Read the report and then answer the questions.

**The paper is by Tom Fahey, Nigel Stocks, Toby Thomas: Quantitative systematic review of randomised controlled trials comparing antibiotic with placebo for acute cough in adults *BMJ* 1998; 316: 906-910.**

### **QUESTIONS ABOUT THIS REPORT:**

- (a) What kind of study is this?
- (b) In the summary we read 'Resolution of cough was not affected by antibiotic treatment (relative risk 0.85 (95% confidence interval 0.73 to 1.00)),' What is a relative risk? What does the 95% confidence interval tell us?
- (c) Why do the authors say that resolution of cough was not affected? Do you agree with this interpretation?
- (d) In the figures, what kind of scale is used for the relative risks? Why are these scales used here?
- (e) In the figures, what do the squares and the horizontal lines represent?
- (f) In the figures, what do the diamond or lozenge shapes represent?
- (g) Do you think the diamonds are drawn correctly?
- (h) In the analysis of side effects of treatment, the authors say: 'When the one trial which reported an increase in side effects from placebo was excluded, the heterogeneity between trials was reduced and side effects were significantly associated with antibiotic use'. What does 'heterogeneity' mean here? Do you think the author's approach is reasonable?
- (i) What are the main conclusions as whether antibiotics should be used in the treatment of acute cough? Are they justified by the data?
- (j) Why would cross-over designs be inappropriate in the study of antibiotic treatment for acute cough?

### **\*\*\* TAKE HOME MESSAGE \*\*\***

Reviewing the literature systematically is time-consuming and must be done thoroughly and meticulously. It is the best way to come to a conclusion about the current state of medical knowledge. The basic principles of statistics are all that is required to understand reviews and meta-analyses, provided they are done properly.

## Research and Critical Skills 15b

7 June 2003

### Examination preparation

*Objectives. To revise some basic ideas of medical statistics and critical reading of the medical literature, in the context of the synoptic examination.*

In the synoptic examination the theme of Research and Critical Skills will be examined in any or all of the following three ways:

1. An extended matching item question on the written paper. This is likely to be attached to one of the clinical scenarios.
2. A component of one of the essays, but the essay may also cover other themes, such as ethics, medicine in society or health and behaviour.
3. A station in the practical examination. This will consist of a published research paper with a set of questions about it. The questions will be answered by one word, a short phrase, a brief sketch, etc. The papers used will be those which we have already covered in the course, i.e. one of those in the Research and Critical Skills handbooks or in the Term 3 Bowel Habit Project, as you could not be expected to get to grips with a new paper in the time available. ***This is an open book question, so you will be allowed to bring your course notes and books with you.***

## Extended matching item questions

*This list covers terms to do with study design, data analysis and interpretation of results. In the examination, EMI questions covering Research and Critical Skills will be integrated with questions on other themes.*

### Item List

#### List A: Study design

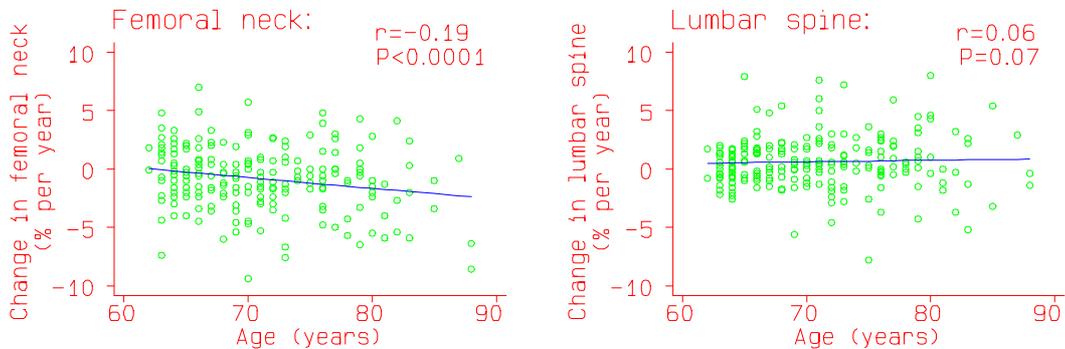
- a Blind assessment
- b Case-control study
- c Clinical case series
- d Clinical case study
- e Cohort study
- f Cross-over trial
- g Cross-sectional study
- h Double blind
- i Matched
- j Prospective
- k Random allocation
- l Random sample
- m Randomized controlled trial
- n Retrospective
- o Sample survey

#### List B: Analysis and interpretation

- a Binomial distribution
- b Chi-squared test
- c Confidence interval
- d Correlation coefficient
- e Large sample comparison of means
- f Large sample comparison of proportions
- g Little or no evidence
- h Normal Distribution
- i Negatively skew
- j Odds ratio
- k Paired t
- l Positively skew
- m Regression
- n Relative risk
- o Scatter diagram
- p Sign test
- q Significance test
- r Standard error
- s Standard deviation
- t Strong evidence
- u Symmetrical
- v Uniform variance
- w Unpaired t
- x Weak evidence

- To study the rates of change in bone mineral density in elderly people, 769 people had bone density measurements made in 1989 and again in 1993. The rates of change in density of the femoral neck and lumbar spine were calculated.

The following graphs show the relationship between rate of change in bone density and age at the initial measurement (male subjects):



(*British Medical Journal* 1994; vol. 309: pages 691-5)

- What term from List A best describes this study design?
  - From List B, what method was used to calculate the lines on the graphs?
  - From List B, what does “r” mean in “ $r = 0.06$ ”?
  - What term from List B best describes the strength of the evidence for the existence of a relationship between rate of change of neck of femur density and age?
  - What term from List B best describes the strength of the evidence for the existence of a relationship between rate of change of lumbar spine density and age?
- In a study of the treatment of frequent, ordinary muscle cramps, subjects were randomly allocated to receive hydroquinone or placebo for two weeks. The subjects were allocated and the tablets packed by investigators who had no contact with patients, and concealed the allocation from the other investigators. The following table was given:

	Hydroquinone	Placebo	
Difference,	group	group	placebo -
hydroquinone	median (inter-	median (inter-	
	quartile range)	quartile range)	(95% CI)
Number of patients	49	53	
Cramps	7 (4-11)	13 (6-18)	5 (2-8)
Cramp-days	5 (3-7)	7 (4-10)	1 (0-3)

(*Lancet* 1997; vol. 349: pages 528-32)

- What term from List A best describes this study design?
- What term from List A best describes the method used to avoid assessment bias in this study?

- (c) What term from List B best describes the shape of the distribution of the number of cramp days?
- (d) What term from List B best describes the strength of the evidence for the existence of an effect of hydroquinone on the number of cramps experienced?
- (e) What term from List B best describes the strength of the evidence for the existence of an effect of hydroquinone on the number of cramp days experienced?

## Essay

- 4. Many asthmatics use inhaled steroids. We want to design a trial to compare an existing inhaler, using a CFC propellant, with a new inhaler with the same drug but a non-CFC propellant. What methodological and ethical considerations would influence the design and conduct of the trial?

## Practical examination question

**Answer the following questions. Only a single word, phrase, sentence, or sketch is required in each case. This is an open book question. You may bring any books to the examination, but you are strongly recommended to bring your course handbooks with your own notes on the papers discussed.**

- 5. **Controlled trial of weight-bearing exercise in older women in relation to bone density and falls (paper in Research and Critical Skills 13b).**
  - (a) What kind of study is this?
  - (b) Is this study double blind?
  - (c) Within what limits do the authors estimate the mean % change in bone mineral content in the ultradistal forearm to lie in older women treated with calcium supplementation only?
  - (d) “We found no significant differences between the groups at baseline.” Why were these tests of significance unnecessary?
  - (e) What confidence intervals would be more useful than the ones in Table 1?
  - (f) The authors state that “The difference between the groups in the number of women falling during the whole two year period was not significant ( $P=0.158$ ) . . .” What method could be used to find “ $P=0.158$ ”?
  - (g) For the difference in proportion of women falling, what analysis would be more informative?