

St. George's Hospital Medical School



Research and Critical Skills

**Common Foundation Programme
Handbook**

2002-3

**Biomedical Sciences
Diagnostic Radiography
Medicine
Nursing
Physiotherapy
Therapeutic Radiography**

J. M. Bland

CFP Research and Critical Skills, 2002-3

Course Handbook

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.

1b. Reading the health care literature -- 1	7
2b. Design of a clinical trial	8
3b. Morphology and Normal Values	9
4b. Summarising and presenting data	12
5b. Significance tests	16
6b. Confidence intervals	19
7b. Reading the health care literature -- 2	21
8b. Revision exercise	23
Specimen In-course Assessment, with Answer	27

In-course Assessments Which You Must Do

In-course Assessment 1, due 30th October	30
In-course Assessment 2, due 27th November	31

NOT INCLUDED HERE:

Course Notes

Glossary of statistical terms

Research and Critical Skills

Timetable: 2002-3, Common Foundation Programme

Code	Date	Time	Place	Topic
Res.01a	Tue 8 Oct	11.30-12.30	Lec	Reading the health care literature -- 1(Prof. Bland)
Res.01b	Thu 10 Oct	10.00-13.30	SG	Critical reading exercise
Res.02a	Tue 15 Oct	10.30-11.30	Lec	Clinical trials (Prof. Bland)
Res.02b	Thu 17 Oct	10.00-13.30	SG	Trials of vitamin C and the common cold
Res.03a	Tue 22 Oct	11.30-12.30	Lec	Optional question and answer session *
Res.03b	Thu 24 Oct	10.00-13.30	Labs	Morphology and normal values: data collection
Res.04a	Tue 29 Oct	11.30-12.30	Lec	Morphology and normal values: summarising data, distributions (Prof. Bland)
Res.04b	Thu 31 Oct	10.00-13.30	SG	Summarising and presenting data.
Res.05a	Tue 5 Nov	11.30-12.30	Lec	Optional question and answer session *
Res.05b	Thu 7 Nov	10.00-13.30	SG	Significance tests
Res.06a	Tue 12 Nov	11.30-12.30	Lec	Optional question and answer session *
Res.06b	Thu 14 Nov	10.00-13.30	SG	Confidence intervals, <u>first assignment due</u>
Res.07a	Tue 19 Nov	11.30-12.30	Lec	Optional question and answer session *
Res.07b	Thu 21 Nov	10.00-13.30	SG	Reading the literature
Res.08a	Tue 26 Nov	10.05-11.05	Lec	MCQ and EMI self test
Res.08b	Thu 28 Nov	10.00-13.30	SG	Revision tutorial, <u>second assignment due</u>

Seminar numbers refer to the chapters in the course book.

Lec: Lecture theatres

Labs: Laboratories, levels 4 & 5 Hunter Wing.

SG: Seminar groups in teaching rooms. Seminar groups will be in three shifts, running in parallel with dissection.

* Optional for those who think they need help. Prof. Bland and/or Dr Peacock will be available to answer students' questions.

St. George's Hospital Medical School
CFP 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

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.
types of data, frequency distribution, histogram, mean, median, quartiles, centiles, range, interquartile range, variance, standard deviation, 95% reference interval or normal range.
rates and proportions, missing denominators, pie charts, 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.
2. Students should be able to:
read a paper from the healthcare literature with critical understanding of basic statistical issues, being aware of problems of design, analysis and interpretation.

Research and Critical Skills, 2002-3

Course Organisation

Course Organizing Team

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

Course structure

This course has four lectures and a self-test session using multiple choice and extended matching items questions. All other 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.

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.

For some sessions, Professor Bland will be available for an hour in the lecture theatre in advance of the seminar should you need help. He will be happy to try to answer any questions you may have, whether they are about the exercise or any other aspect of the course.

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 Prof. Bland (mbland@sghms.ac.uk, ext. 5492, room 6.02 Hunter Wing).

Marked assignments

Twice during the term you will be set an assignment, which will consist of an appraisal of a paper from the healthcare literature. The first assessment will be formative, the second will be summative and will form part of the end of term assessment. The first assignment will be marked and returned, with a specimen answer, before the second is due in. Both of these assignments are included at the end of this handbook. The dates when they are requested are in the timetable.

Solutions to the exercises

You will be given specimen answers for the marked assignments and the multiple choice tests. 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.

Examples used

All the examples used in the seminar exercises and course notes are genuine and references to their source in the health care literature are given.

Attendance

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

Recommended books and course notes

Course notes are included in this handbook. 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.

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

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

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

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

Leaveron 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.

Two books of exercises should be helpful:

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

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.)

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 Computing Resource Rooms, and call up Applications on the Intranet. Network Services. Click "Statistics" 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 resource rooms have SPSS.

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

Methods of Assessment

All students will be assessed using an in-course assessment. This will be an exercise in reading a published paper. There will be two assessments. Both of these will be marked and specimen answers given. For most courses the second assessment forms part of the mark for the term.

All students will do a written exam at the end of Term 1, including either MCQ or EMI questions in Research and Critical Skills. There will be one formative test and revision session using these types of question, which will be marked by yourself from a sheet of solutions. Further practice questions and their answers can be found on the course web page.

Medicine students will do a synoptic exam at the end of Term 3 which will include the CFP material. 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 paper reading exercise will be open-book and will ask questions about one of the research papers which we shall be reading in the course.** Attendance at small-group sessions is therefore strongly advised, as you will be able to bring your notes into the practical exam with you.

Research and Critical Skills 1b

10 October 2002

Reading the Health Care Literature -- 1

Read Course Notes Chapter 1 before this exercise.

Objectives: to read critically a paper from the medical literature.

This is a short paper from the *British Medical Journal* for 18th September 1999. (This article was available in full on the *BMJ* web site, whence it has been downloaded and printed here.) It is a short report, which has an abbreviated structure compared to a full length paper.

Split into small groups of 4 or 5. Read the paper and then look at the questions below. Decide on your small group's answers. These will then be compared.

The paper is S G M Luke, A Gallagher, B W Lloyd, Staff and family attitudes to keeping joint medical and nursing notes at the foot of the bed: questionnaire survey *BMJ* 1999; 319: 735-735.

Questions about this report

1. How does this paper match the usual structure of Summary, Introduction, Methods, Results, Discussion, Conclusions?
2. What is the purpose of the study? Is this clearly explained?
3. How was the study carried out? Is there sufficient information for another researcher to repeat the study?
4. What did they actually observe?
5. Do they mention any limitations of their study?
6. What are the authors' conclusions? Do you agree with them?

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

Medical research provides essential information for decision making, but it cannot be accepted at face value. It must be read carefully and critically.

Research and Critical Skills 2b

17 October 2002

Design of a clinical trial

Read Course Notes Chapter 2 before this exercise.

Objectives. At the end of this exercise you 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.

In this exercise we shall look at some of the practical problems involved in designing a research study, in this case a clinical trial.

The class will be divided into two subgroups. Each subgroup will consider a separate trial. After deciding on a plan for the trial, each group should choose one or more spokespersons to present an account of their study design and the thinking behind it. This is then open for comment by the rest of the class.

There is no single right answer. Several different study designs are possible, and several different studies have been done. Each trial is concerned with the effects of 1g vitamin C tablets on the common cold. The first concerns prevention, the second treatment.

Group 1: Does taking 1g vitamin C daily reduce the risk of getting a cold?

Group 2: Does taking 1g vitamin C daily after a cold begins alleviate the symptoms of a cold?

Points you should consider

Choice of sample. What sort of subjects do you want? How will you get them? How will you persuade them to cooperate?

Informed consent. How much will you tell your subjects about the trial?

Choice of control treatment. Do you need a placebo? Do you need to give any special instructions about other sources of vitamin C?

Allocation to treatments. How done? Two sample or crossover design?

Outcome variable. What is your outcome variable? How are you going to measure it? Self reports? Clinical examination? Measurements? Should your assessment be blind and how do you achieve this?

Other factors. Are there any other factors that should be taken into account? What about other treatments or prophylactics?

Conclusions. What observations would convince you of a treatment effect?

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

The best way to test a new treatment is by a randomized clinical trial. This ensures groups are comparable before treatment. If possible this should be double-blind, i.e. neither experimenter nor subject knows which treatment is given. This ensures that there is no response bias due to the knowledge of being treated or to the circumstances of the study, and no assessment bias due to the experimenters favouring one treatment over another. In practice this can be difficult to achieve.

(I am greatly indebted to Stuart Pocock, on whose idea this exercise is based.)

Research and Critical Skills 3b

24 October 2002

Morphology and Normal Values

Read Course Notes Chapter 3 before this exercise.

Objectives: to see how body size varies among normal individuals, how measurements on the same subject can vary, how statistical methods including histograms, means, and standard deviations, can be used to describe normal values and how these might be used in the diagnosis of disease.

In this practical you will make some measurements and observations on one another: height, upper arm and head circumference, pulse, eye colour, and sex. Each will be observed twice for each student.

Equipment required: stadiometer, tape measure, watch (use your own).

In groups of up to 10 students, choose two scribes from the group. The scribes will be responsible for recording the data and entering them onto a computer file via the Intranet.

Form pairs. Each member should measure the following, once, on the other, noting the results:

1. Height in mm, using the stadiometer. This should be without shoes.
2. Upper arm circumference in mm, using the tape measure. The tape should be put firmly but not tightly round the middle of the right upper arm.
3. Head circumference in mm, using the tape measure. The tape should be put firmly but not tightly round the head at the level of the forehead.
4. Pulse rate. Find the pulse on the left inner wrist. Time for 60 seconds while counting the beats.
5. Eye colour. Make sure you observe this. Do NOT ask the subject what they think. Record the colour of the iris as one of the following:
 1. black
 2. brown
 3. blue
 4. grey
 5. hazel
 6. green
 7. other
6. Sex. Report as
 1. female
 2. male

Pass the measurements to the scribes, who should record them on the form provided. The subject's name is to ensure that the first and second sets of data are recorded for the same student. It will **not** be entered into the computer.

Now make the second set of measurements. Form new pairs and repeat the measurements. Pass them to the scribes, who should record them on the form provided.

When all measurements are complete, the appointed scribes should take the form to the computer resource centre. The data should be entered into a computer via the intranet. Go to the Research and Critical Skills page and use the data entry program which has been set up for this practical.

The combined measurements of all students will then be presented in the lecture theatre on Tuesday 29th October.

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

The size and shape of bodies vary. We need to know how the observations for normal people behave before we can interpret observations on patients. The distribution can be summarized by statistics such as mean and standard deviation. About 2/3 of observations will be within one standard deviation from the mean and about 95% will be within two standard deviations. We can calculate the 95% reference interval from these, but may have to transform the data first to make the distribution symmetrical. Most biological variables cannot be measured without error.

Research and Critical Skills 4b

31 October 2002

Summarising and Presenting Data

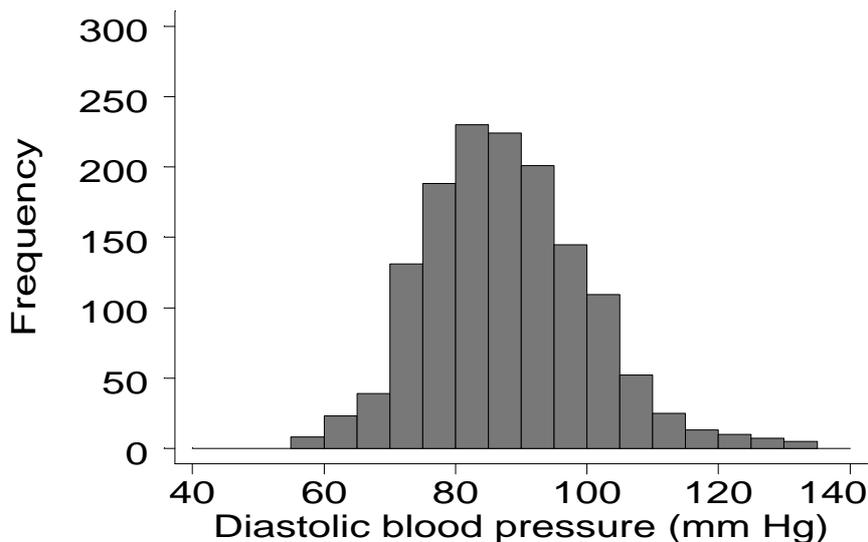
Read Course Notes Chapter 3 before this exercise.

Objectives. At the end of this exercise and after reading Chapter 3 of the Course Notes, you should be able to recognise, describe, and know the appropriate uses of: histograms, medians and other quartiles and centiles, range and inter-quartile range, mean and standard deviation, and normal or reference intervals.

We want you to answer the questions, using the course 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:

In a study of blood pressure in one town, the distribution of diastolic blood pressure among men was shown as follows (*British Medical Journal*·1974; **3**: 600-3):



- How would you describe the shape of the distribution?
- The class intervals are 5mm Hg wide. In what interval would a diastolic blood pressure of 70 be put?

In a study of muscle training in patients with rheumatoid arthritis, patients were randomized to receive either a standardized exercise regime or normal care. Measurements were made before and after six weeks' treatment. The observer was blind to the treatment (*Physiotherapy Research International* **4**: 55-67).

The following tables were given among the results:

TABLE 1: The study population

Characteristic	Exercise group	Control group
Number	17	18
Females (males)	15 (2)	14 (4)
Age (mean \pm SD) (years)	51.4 (\pm 11.1)	49.7 (\pm 15.3)
Body weight (mean \pm SD) (kg)	71.0 (\pm 23.4)	69.4 (\pm 17.9)
Height (mean \pm SD) (m)	1.63 (\pm 0.6)	1.66 (\pm 0.8)

TABLE 2: Mean (\pm SD) for parameters measured in the control and experimental subjects before and after completion of the six-week study

Measure	Control group		Experimental group	
	Pre-test	Post-test	Pre-test	Post-test
Peak speed (deg/sec ⁻¹)	125.2 (\pm 33.9)	121.6 (\pm 27.7)	132.0 (\pm 40.6)	154.0 (\pm 45.8)
TUG (sec)	12.6 (\pm 1.5)	12.2 (\pm 1.3)	11.7 (\pm 1.9)	10.4 (\pm 1.8)
P _{VAS} (mm)	4.1 (\pm 2.4)	3.9 (\pm 2.4)	4.3 (\pm 2.2)	2.4 (\pm 2.1)
HAQ	0.7 (\pm 0.4)	0.8 (\pm 0.3)	1.0 (\pm 0.6)	0.7 (\pm 0.6)

In these tables, "post-test" means the test done after the six weeks of treatment. Peak speed is the peak angular velocity of the knee movement in rising from a sitting position, measured in degrees per second. TUG is the time to get up and go from a sitting position, in seconds. P_{VAS} is pain measured on a visual analogue scale, where patients are given a 10 cm line, marked "no pain at all" at one end and "worst pain you can imagine" at the other. The patient marks the point which they think best represents the pain. HAQ, health assessment questionnaire, is a scale of limitations on activity, high scores being more limited. (*Physiotherapy Research International*, 1999; 4: 55-67)

- (c) What does "SD" stand for and what does it mean?
- (d) In Table 1 there are two obvious misprints. What are they?
- (e) In Table 2, there are two obvious misprints in the units in which variables are measured. What are they?
- (f) In Table 2, which variables must have a skew distribution?

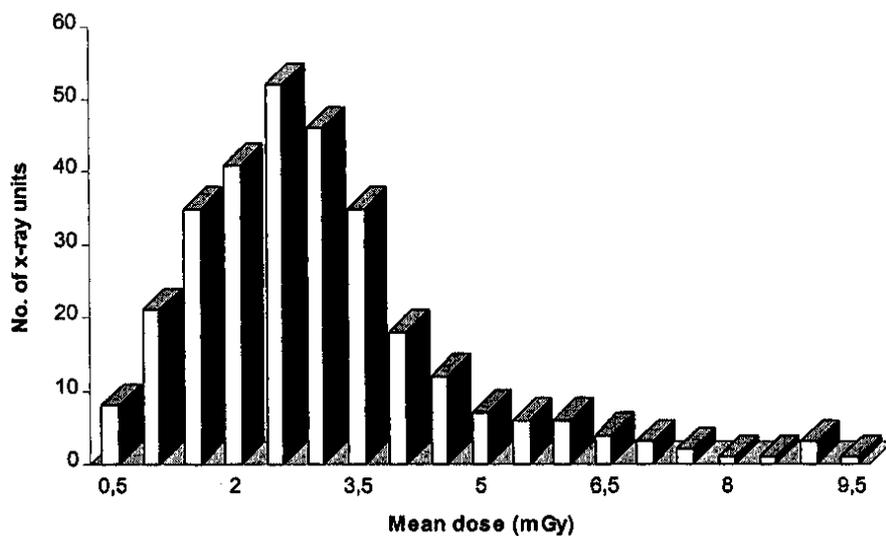
The normal, reference or 95% range for haematocrit in men is 0.39 to 0.55 (*Geigy Scientific Tables Vol. 3*).

- (g) What is meant by this statement?
- (h) What would it tell us about a man with a haematocrit of 0.47?
- (i) What would it tell us about a man with a haematocrit of 0.38?
- (j) What would it tell us about a woman with a haematocrit of 0.38?

Subgroup 2:

A study was carried out of the radiation dose used in dental X-ray diagnostics, based on quality control tests performed on 307 X-ray odontology installations in Spain. (*British Journal of Radiology*, 2001; **74**, 153-156.) The estimated mean exposure value was 2.89 ± 2.12 mGy, with a median of 2.43 mGy, and the 75th percentile was 3.37 mGy. The following figure shows a frequency plot, which excludes five cases with mean dose values above 10 mGy.

- The authors report that the does had "median of 2.43 mGy, and the 75th percentile was 3.37 mGy." What does this statement mean? What is the relationship between the third quartile and the 75th percentile?
- Figure 2 is described in the paper as a "frequency plot". What other name is usually given to such a diagram? How could the presentation of this graph be improved?



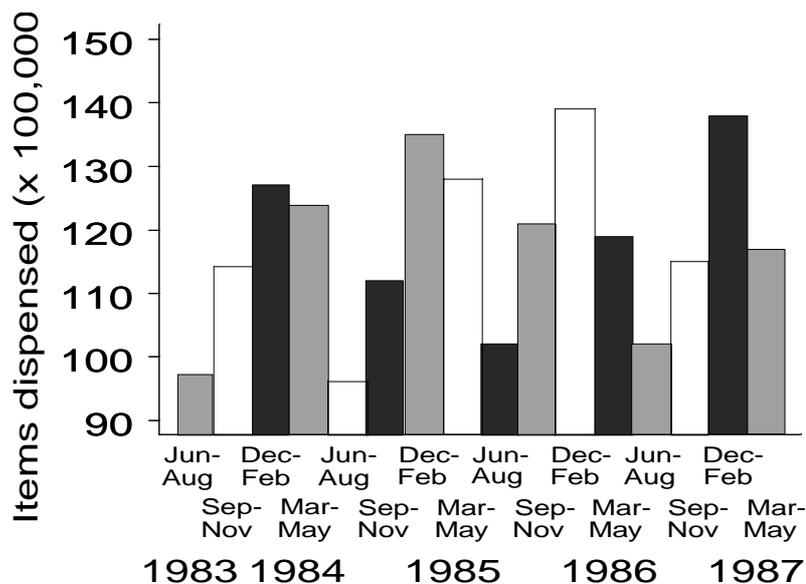
- What term would be used to describe the shape of this frequency distribution?
 - Five points have been omitted. What would be the effect on the distribution of including them?
 - The mean dose was 2.89 with standard deviation 2.12 mGy. Why did the authors say "The standard deviation suggests that very few X-ray units irradiate the patient above 7 mGy (less than 5% based on this work)"?
2. In a study of patients admitted to an otolaryngology ward, 140 patients with nose-bleeds were compared to 113 controls with other conditions. Patients were interviewed about their alcohol consumption (*British Medical Journal* 1994; **309**: 640). The results for the number of units of alcohol consumed per week were:

	Nose-bleed patients (n=140)	Other patients (n=113)
Median	10.0	2.0
Inter-quartile range	0 to 50	0 to 10

(One unit of alcohol is 1/2 pint beer, a glass of wine, or a pub measure of spirits.)

- (f) What is meant by "inter-quartile range"? Why are the median and inter-quartile range used here, rather than mean and standard deviation?
- (g) What is the shape of the distribution of alcohol consumption?

3. Many prescribed drugs show a seasonal pattern in their use. The following graph shows an example of this and is taken from routinely collected Prescribing Analyses and Cost (PACT) data (*Practitioner*, 1991; **235**, p450-1):



- (h) What kind of graph is this?
- (i) What feature of the graph might make it misleading?
- (j) What feature of the graph makes it difficult to see the pattern?
- (k) How could the graph be improved?

***** TAKE-HOME MESSAGE *****

The standard deviation is a very useful statistic, which can be used to summarise data and tell us something about the shape of the distribution and to calculate reference ranges.

Research and Critical Skills 5b

7 November 2002

Significance tests

Read Course Notes Chapters 4 and 5 before this exercise.

Objectives: To understand the application of probability as used in significance tests in health care research.

Subgroup 1

Read the attached paper, Eva Pettersson, Ann Gardulf, Gun Nordström, Christina Svanberg-Johnsson, Gunnar Bylin. Evaluation of a nurse-run asthma school. *International Journal of Nursing Studies* 1999; **36**: 145-151
uscle. *Journal of Cardiac Failure* 2. 117-117.

Questions about this paper

- From Table 2, what can we conclude about the shape of the distribution of the overall Sickness Impact Profile? What three features of the data support this?
- In Section 3.1, what is meant by " $p < 0.001$ "? What can we conclude from this?
- In Section 3.2, what is meant by "n.s."? What can we conclude from this?
- What limitations does the lack of a control group lead to?
- What bias, if any, might there be in the patients' response concerning the use of a PEF-meter (Section 3.3.)?
- What bias, if any, might there be in the FEV1 measurement (Section 3.4.)?

Sub-group 2

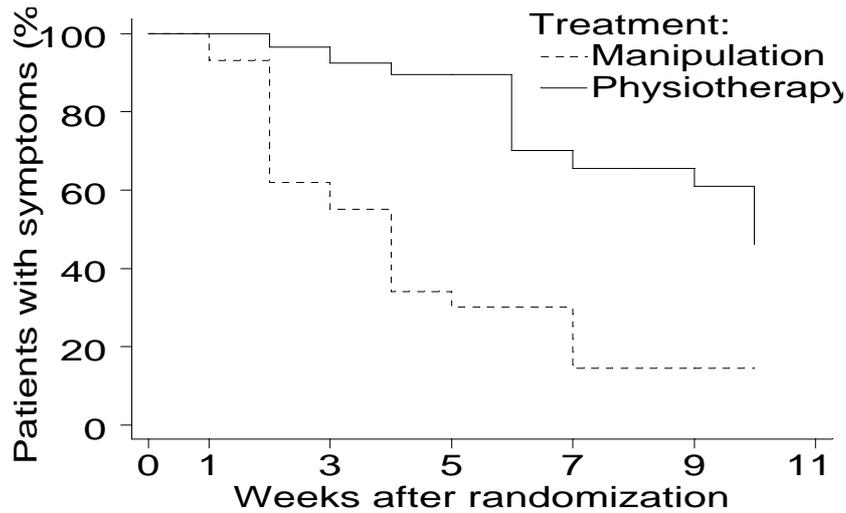
The following is extracted from a paper (*British Medical Journal* 1997; **314**: 1320-5) entitled: "Comparison of physiotherapy, manipulation, and corticosteroid injection for treating shoulder complaints in general practice: randomised, single blind study", by Winters, J.C., Sobel, J.S., Groenier, K.H., Arendzen, H.J., and Meijboom-de Jong, B. The questions are embedded in the paper.

Winters *et al.* carried out a single-blind randomized trial comparing physiotherapy, manipulation and corticosteroid injection for treating shoulder complaints in general practice. The subjects were divided into two diagnostic groups: a shoulder girdle group ($n = 58$) and a synovial group ($n = 114$). Patients in the shoulder girdle group were randomized to manipulation or physiotherapy, and patients in the synovial group were randomized to corticosteroid injection, manipulation, or physiotherapy. Corticosteroid injections were given by the participating doctors immediately after randomization, one week later, and, if needed, after a further two weeks. Manipulation was done by either the participating general practitioners or physiotherapists. They were instructed in which techniques to use. Physiotherapy was given twice a week by local physiotherapists. No mobilisation techniques or manipulative techniques were allowed, to keep the manipulation and physiotherapy regimes distinct. After treatment had started, the patients were asked to complete a pain questionnaire each week. This gave a pain score, obtained by rating seven aspects of pain, each from one to four, giving a total score between 7 points (no pain) to 28 points (severe

pain). They were also asked to indicate if they felt cured or if they thought that the treatment had failed. Follow-up physical examinations after randomization were done by a physiotherapist who was not informed about the patients' diagnosis and treatment.

- (a) This is described as a single blind study. Who is blinded and why was this done?
- (b) How easy would it be to break the blinding?
- (c) What might be the effects of the problems with blinding in this study?

The figure below shows the percentage of shoulder girdle patients still with symptoms at different times after the start of the trial.



- (d) With reference to this graph, the authors say that "manipulation was superior to physiotherapy ($P < 0.001$)". What does " $P < 0.001$ " mean?
- (e) What can we conclude from this test?

The pain scores (high score implies more pain) of patients in the shoulder girdle group, at randomization to treatment and at the end of treatment (when the patient left the study or 11 weeks after randomization) were compared. The following table shows the mean (SD) of the pain scores at each time:

	Manipulation (n=32)	Physiotherapy (n=35)
At randomization	14.8 (4.2)	14.4 (3.5)
At end of treatment	9.9 (3.5)†	12.0 (4.4)†

† Significant difference from pain score at randomization ($P < 0.001$)

- (f) What can we conclude from these P values?
- (g) Why is this not the most useful comparison which could be made?
- (h) What comparison would have been more useful?

For the synovial group, analysis of the time spent in the symptomatic state showed a significant difference between the treatment groups ($P < 0.001$). The corticosteroid injection group improved rapidly, while the physiotherapy group improved slowly and the manipulation group did only slightly better. At five weeks after randomization, 75% of patients in the injection group were "cured" compared with 20% in the physiotherapy group and 40% in the manipulation group. Drop-out because of treatment failure was much lower in the injection group (17% (7/47)) than in the physiotherapy group (51% (18/35)) and manipulation group (59% (19/32)).

Some patients were reported to have been "cured" before 11 weeks after randomization. Among these patients, a recurrence of complaints by week 11 was reported by 18% (7/39) of patients in the injection group, 13% (2/15) in the physiotherapy group, and 8% (1/13) in the manipulation group.

- (i) What difficulty is there in interpreting these recurrence data?

The authors concluded that the results suggest that manipulation is to be preferred to physiotherapy for treating shoulder complaints originating from the shoulder girdle in general practice. They also conclude that injection is the most effective treatment for shoulder complaints originating from the synovial structures in general practice.

- (j) On the basis of the partial summary given here, do you think that these conclusions are reasonable?

Research and Critical Skills 6b

14 November 2002

Confidence intervals

Read Course Notes Chapters 7 and 8 before this exercise.

Objectives. At the end of this exercise you should be able to recognise, describe, and know the appropriate uses of confidence intervals.

We want you to answer the following 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 two short papers here appeared together in the British Medical Journal. They are very similar in the question they address and the methods used. Both refer to "moxibustion". In moxibustion, moxa, a preparation of wormwood (*Artemisia moxa*), or another slow-burning substance, is lit and held as near to the point on the skin as possible without causing pain or burning. Both also refer to "significant" events. The word is not used in its statistical sense.

Sub-group 1:

Read the paper by Adrian White, Simon Hayhoe, Anna Hart, Edzard Ernst: Adverse events following acupuncture: prospective survey of 32 000 consultations with doctors and physiotherapists *BMJ* 2001;323:485-486.

Questions about this report

- (a) Are there any problems with the sampling method? What alternative methods might have been used? Would they solve the problem?
- (b) Are there any problems with the data collection methods? What alternatives might have been used? Would they solve the problem?
- (c) The average age of the acupuncturists was 47 (range 27-71) years. The median number of consultations for a practitioner was 318, range 5-1,911. What does this tell us about the shapes of the distributions of age and number of consultations?
- (d) Altogether, 43 "significant" events were reported, giving a rate of 14 per 10,000 (95% confidence interval 8/10,000 to 20/10,000). What does this mean?
- (e) According to accepted criteria, none (0/10,000 to 1.2/10,000) of these events was serious. Can we conclude that there is no risk of serious events?
- (f) The authors say "14 per 10,000 of these minor events were reported as significant. These event rates are per consultation, and they do not give the risk per individual patient". Why do they not give the risk per individual patient?
- (g) The authors do not appear to draw any explicit conclusions. What would you conclude from this study?

Sub-group 2:

Read the paper by Hugh MacPherson, Kate Thomas, Stephen Walters, Mike Fitter:
The York acupuncture safety study: prospective survey of 34 000 treatments by traditional acupuncturists. *BMJ* 2001;323:486-487.

Questions about this report

- (a) Are there any problems with the sampling method?
- (b) What problems might the low response rate from the acupuncturists lead to?
- (c) Are there any problems with the data collection methods? What alternatives could be used? Would they solve the problem?
- (d) The mean age of participants was 44.8 years (range 23-79 years). What does this tell us about the shape of the distribution of age? Would we expect the median age to be less than or greater than 44.8 years?
- (e) Practitioners reported 43 minor adverse events, a rate of "1.3 (0.9 to 1.7) per 1,000 treatments". What is "(0.9 to 1.7)" and what does it tell us?
- (f) The authors conclude that their data are consistent with an underlying serious adverse event rate of between 0 and 1.1 per 10,000 treatments. Is this a reasonable interpretation?
- (g) The authors say that further research measuring patients' experience of adverse events is merited. What would this tell us that these papers do not?

Research and Critical Skills 7b

21 November 2002

Reading the health care literature -- 2

Read Course Notes Chapter 6 before this exercise.

Objectives. At the end of this exercise you should have gained experience in critically reading a paper from the health care literature.

Subgroup 1

Read the paper by M Clare Robertson, Nancy Devlin, Melinda M Gardner, A John Campbell: Effectiveness of a nurse delivered home exercise programme to prevent falls. *British Medical Journal* 2001; 322: 697-701.

Questions about this paper:

- (a) What does "randomised" mean? Why were subjects randomised to the exercise programme or to usual care?
- (b) Written informed consent was obtained. What does this mean and why is it important?
- (c) The "mean (SD)" of the number of current prescribed drugs were 3.1 (2.4) and 2.9 (2.3) for the two groups. What does this mean? What does it tell us about the shape of the distribution of the number of prescribed drugs taken?
- (d) "We found a 46% reduction in the number of falls during the trial for the exercise group compared with the control group (incidence rate ratio 0.54, 95% confidence interval 0.32 to 0.90)." What is a 95% confidence interval? What does this tell us about the difference in incidence of falls?
- (e) What might be the effect on the interpretation of the results of some subjects not completing the trial?
- (f) There were 9 serious falls in the control group and 2 in the exercise group, $P=0.033$. What is meant by " $P=0.033$ "? What does it tell us about the effect of exercise on falls?
- (g) In Table 2, P values are only given when the difference is statistically significant. The difference for injurious falls is therefore not statistically significant. What can we conclude from this?
- (h) The main conclusion of the paper is "An individually tailored exercise programme delivered at home can prevent falls." Do the data support this?

Subgroup 2

Read the paper by M Garrett, A M McElroy, A Staines: Locomotor milestones and babywalkers: cross sectional study. *British Medical Journal* 2002;324:1494.

Questions

- (a) Why is this described as a cross sectional study?
- (b) How does this design differ from a randomised trial? How does this affect the interpretation of the results?
- (c) To what problems might this method of data collection have led?
- (d) Baby walker use was reported to start at a median age of 26 weeks, (interquartile range 26-28). What does this tell us about the distribution of age of first use?
- (e) In the table, the difference in mean time to stand alone was 3.32 weeks, (95% confidence interval 4.87 to 1.77). What does this statement mean and what can we conclude?
- (f) In the table, the P value for the difference in mean time to stand alone was <0.0001. What does this statement mean and what can we conclude?
- (g) Can we conclude that babywalkers delay development?
- (h) How might the data on time to roll over influence our interpretation of the results of this study?

Research and Critical Skills 8b

28 November 2002

Revision exercise

There is no new material for this exercise.

Objectives. To gain practice in explaining statistical terms used in the medical literature. To revise statistical concepts and consolidate knowledge.

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.

Sub-group 1:

The following statement appeared in a letter from a British Labour Party spin-doctor in a national newspaper: "The average income of the middle quartile of earners (40% to 60% of the population) is £326 a week . . ." (Gould 1999).

- (a) What is wrong with this statement?
- (b) What is the more usual name for the "middle quartile"?

Two treatments used to stop breast milk secretion were compared. One was bromocriptine twice daily for fourteen days, the other was a single dose of cabergoline. Two hundred and seventy two women were randomized to two equal sized groups. One group received a placebo after delivery and two doses of bromocriptine each day for 14 days, the other received a dose of cabergoline after delivery and a placebo twice daily for 14 days. Complete success in inhibiting lactation was obtained in 69% of the bromocriptine group and 78% of the cabergoline group. The difference was not significant and the 95% confidence interval for the difference was 19 to 2 percentage points. (*BMJ* 1991: 302;1367-1371)

- (c) What is meant by "randomized"? Why was this done?
- (d) Why were placebos used and why was each woman given a placebo treatment?
- (e) What is meant by "not significant"?
- (f) What can be concluded about the relative efficacy of the treatments?

An M.Sc. dissertation looked at the average time taken to complete a task. The following table was produced:

	Mean time (days)	95% confidence interval
Intervention group	1.74	0.19 to 1.97
Comparison group	2.34	1.13 to 0.39

(g) What features of this table suggest that something is wrong?

Sub-group 2:

The following extract is from *Hansard*, 29 November 1991 (quoted in *Royal Statistical Society News and Notes*, 1992 vol. 18(7) page 12). The idea which Members of Parliament are discussing is that a minimum wage would be defined as a fixed proportion of the median wage.

Mr Arbuthnot: . . . suggestion of a minimum wage is in itself rather obscure and bizarre. As I understand it, it is tied to the average and would therefore not only be relatively high at £3.40, but would increase as the average wage itself increased. With each increase in the average rate of pay, the minimum wage itself would have to go up and it would be forever chasing its own tail.

Mr Tony Lloyd: Perhaps I can help the Hon. Gentleman. It will be tied to the median, which is not the same as the average. It is simply the mid-point on the range and would not be affected by changes in the minimum wage.

Mr Arbuthnot: From what I understand, even an amount tied to the median would be affected because if the lowest wage were increased to £3.40 per hour, the median would have to rise.

Mr Tony Lloyd: I shall put the matter in simple terms. The median, the mid-point in a series of numbers such as 2, 2, 5, 6 and 7, is defined as being the difference between 2 and 7, which is 3.5. If we alter the figures 2 and 2 to 3.5, the middle figure of 5 would remain unaltered because it is independent of the bottom figures.

Mr Arbuthnot: I do not understand the Hon. Gentleman's mathematics and I slightly doubt whether he does.

Mr Matthew Carrington: I am extremely confused. I studied mathematics for some years at school and I have not totally forgotten all of them. The median is not the mid-point between the first number and the last. It is where the largest number of items in a sample comes to, whereas the average is obviously the sample multiplied by the number of items. The Hon. Member for Stretford (Mr Lloyd) is obviously extremely confused. The median has a precise mathematical definition which is absolutely right, and my Hon. Friend is correct in saying that the median is bound to alter if the number at the bottom of the scale is changed. That will alter the average as well in a different way, but it is bound to alter the median. Perhaps the Hon. Member for Stretford wishes to define median in a non-mathematical sense.

Mr Arbuthnot: I am extremely grateful to my Hon. Friend for sorting out at least the Hon. Gentleman's mathematics with obvious skill and knowledge.

- Which Honourable Member is correct, if any, and why?
- What would be the effect on the skewness of the earnings distribution if the minimum wage were made a fixed proportion of the median, assuming that this figure was then higher than the current wage of some members of the population?

The following is the abstract of a paper reporting a clinical trial (*BMJ* 2001; **322**: 261-265):

Objective: To compare the effectiveness of nasal diamorphine spray with intramuscular morphine for analgesia in children and teenagers with acute pain due to a clinical fracture, and to describe the safety profile of the spray.

Design: Multicentre randomised controlled trial.

Setting: Emergency departments in eight UK hospitals.

Participants: Patients aged between 3 and 16 years presenting with a clinical fracture of an upper or lower limb.

Main outcome measures: Patients' reported pain using the Wong Baker face pain scale, ratings of reaction to treatment of the patients and acceptability of treatment by staff and parents, and adverse events.

Results: 404 eligible patients completed the trial (204 patients given nasal diamorphine spray and 200 given intramuscular morphine). Onset of pain relief was faster in the spray group than in the intramuscular group, with lower pain scores in the spray group at 5, 10, and 20 minutes after treatment but no difference between the groups after 30 minutes. 80% of patients given the spray showed no obvious discomfort compared with 9% given intramuscular morphine (difference 71%, 95% confidence interval 65% to 78%). Treatment administration was judged acceptable by staff and parents, respectively, for 98% (199 of 203) and 97% (186 of 192) of patients in the spray group compared with 32% (64 of 199) and 72% (142 of 197) in the intramuscular group. No serious adverse events occurred in the spray group, and the frequencies of all adverse events were similar in both groups (spray 24.1% v intramuscular morphine 18.5%; difference 5.6%, 2.3% to 13.6%).

Conclusion: Nasal diamorphine spray should be the preferred method of pain relief in children and teenagers presenting to emergency departments in acute pain with clinical fractures. The diamorphine spray should be used in place of intramuscular morphine.

(Diamorphine hydrochloride is heroin.)

- (c) No blinding was used in this study. What effect might this have?
- (d) Written informed consent was obtained from a parent or guardian. Oral consent was also obtained from the patient if aged over seven years. Why was this necessary? What should be done if the patient does not consent but the parents do?
- (e) When consent had been obtained and inclusion and exclusion criteria met, the next numbered case report form was opened. Randomised allocation codes, prepared before the start of the study by BCR, were concealed in sealed opaque envelopes in the case report form. Why was it important that the envelopes be opaque?

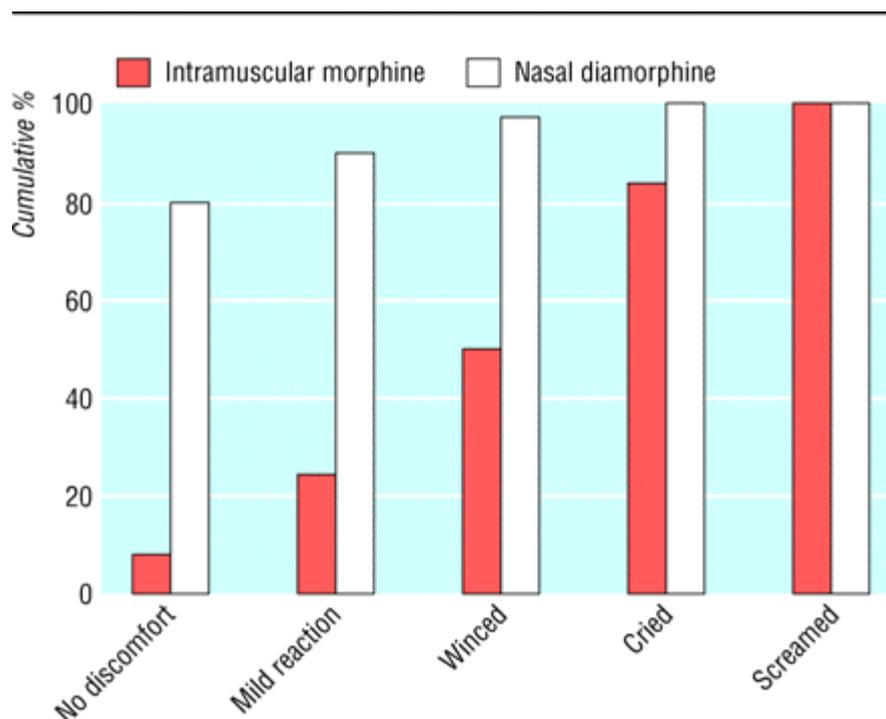
- (f) The authors stated that "The distribution of pain scores for the spray group was lower than that for the intramuscular group at 5 (P=0.04), 10 (P=0.003), and 20 minutes (P=0.002) after treatment, but no different after 30 minutes (P=0.20)". What can we conclude about pain score after 10 minutes? The pain scores after 30 minutes were as follows:

Pain score	Nasal diamorphine group n (%)	Intramuscular morphine group n (%)
1	54 (28)	48 (25)
2	65 (34)	61 (32)
3	36 (19)	42 (22)
4	24 (13)	23 (12)
5	7 (4)	11 (6)
6	5 (3)	8 (4)

Is it correct to conclude that pain scores were no different after 30 minutes?

- (g) What is meant by "confidence interval 65% to 78%"? What can we conclude about the difference?

The following graph was used to present the data on reaction to the treatment:



- (h) What does this graph actually show? How could it be improved?

***** TAKE-HOME MESSAGE *****

There is no particular take-home message for this session.

Research and Critical Skills

Specimen In-course Assessment

This is was the summative assessment for 2001-2, which contributed to the total mark for this term. It is included here to show what sort of answers we are looking for in the in-course assessments. A specimen answer follows.

Read the paper by Didier Pittet, Stéphane Hugonnet, Stephan Harbarth, Philippe Mourouga, Valérie Sauvan, Sylvie Touveneau, Thomas V Perneger, and members of the Infection Control Programme: **Effectiveness of a hospital-wide programme to improve compliance with hand hygiene.** (*Lancet*, 2000, Volume 356, pages 1307-12) and answer the questions. This paper has not been edited and contains a number of statistical terms beyond the scope of the CFP course. You do not need to know these to understand the paper or answer the questions. These definitions might help. MRSA (Methicillin resistant *Staphylococcus aureus*) is an infection which is resistant to many antibiotics. Nosocomial infections are those acquired in hospital. The Hawthorne effect is the process by which awareness of being studied alters subjects' behaviour.

Answer the following questions. You should only need 2 to 6 sentences per answer.

- (a) "Overall compliance improved from 47.6% in 1994, to 66.2% in December 1997 ($p < 0.001$; figure 1)." What does " $p < 0.001$ " mean here and what can we conclude from it?
- (b) "Although not statistically significant, similar trends were observed in gynaecology/obstetrics ($p = 0.17$), and paediatric wards ($p = 0.12$; figure 2A)." What does "not statistically significant" mean? What can we conclude about the association between handwashing campaigns and compliance in gynaecology/obstetrics and paediatric wards?
- (c) How strong is the evidence that nosocomial infections and MRSA showed a systematic decline over time?
- (d) In December 1994, the compliance was 47.6%, 95% CI 46.8—48.5 (Table 2). What does 95% CI 46.8—48.5 mean and what does it tell us?
- (e) As the authors say, this was not a randomised trial. To what problems does the lack of a comparable control group lead?
- (f) The authors say that "The ethical acceptability of control groups in situations perceived as threatening to patients (high endemic nosocomial infection and MRSA transmission rates) was an additional obstacle [to randomisation]". What arguments could be put against this view?

Specimen Answer to In-course Assessment 2

This is a possible answer to the exercise. It does not necessarily cover all possible points in the answers, but it is the sort of thing we were looking for. Some of the answers are rather longer than we asked for, because I tried to include all relevant points.

Your script has been retained so that the external examiners can see it if they wish. The mark should be posted by the end of term.

- (a) *"Overall compliance improved from 47.6% in 1994, to 66.2% in December 1997 ($p < 0.001$; figure 1)." What does " $p < 0.001$ " mean here and what can we conclude from it?* This is the result of a significance test. It is testing the null hypothesis that, in the population which these hospital workers represent, compliance has not changed between 1994 and 1997. If the null hypothesis were true, the probability of observing a difference as large as this in a sample is less than 0.001. The difference is highly significant and we can conclude that there is very strong evidence that compliance has increased. The observed difference is not consistent with the null hypothesis.
- (b) *"Although not statistically significant, similar trends were observed in gynaecology/obstetrics ($p = 0.17$), and paediatric wards ($p = 0.12$; figure 2A)." What does "not statistically significant" mean? What can we conclude about the association between handwashing campaigns and compliance in gynaecology/obstetrics and paediatric wards?* "Not statistically significant" means that the probability of observing so large a difference if the null hypothesis were true would be greater than 0.05. We can conclude that there is little or no evidence that compliance has increased on wards for these specialties. The observed differences are consistent with the null hypotheses. We cannot conclude that compliance has not increased, only that we have failed to detect an increase.
- (c) *How strong is the evidence that nosocomial infections and MRSA showed a systematic decline over time?* Nosocomial infections increased in 1995, declined sharply in 1996 and 1997, then rose slightly. MRSA appeared to increase from 1993 to 1994 (before the campaign), then declined steadily for 4 years, then rose slightly (Figure 3). Between 1994 and 1998, the rates of each halved, which is a big change. Comparing the proportions in 1994 and 1998, these differences were significant ($P = 0.04$ and $P = 0.02$) providing some evidence that these infections fell. The P values for MRSA rates per 10000 patient days are smaller, < 0.001 , suggesting that there is strong evidence for a decline.
- (d) *In December 1994, the compliance was 47.6%, 95% CI 46.8—48.5 (Table 2). What does 95% CI 46.8—48.5 mean and what does it tell us?* This is a 95% confidence interval. It is a range of possible values within which we estimate that the compliance in the whole population lies. It is chosen so that for 95% of possible samples, the confidence interval would include the population compliance. It tells us that we can estimate the compliance in the population to be between 46.8% and 48.5%. The confidence interval is quite narrow, suggesting that the estimate is reasonably accurate.
- (e) *As the authors say, this was not a randomised trial. To what problems does the lack of a comparable control group lead?* We do not know whether any changes which took place over the study period would have taken place anyway, without the educational campaign. Staff would be exposed to other influences apart from the campaign. A control group, of

staff who did not experience the campaign, would make it easier for us to conclude that any differences were the effect of the educational campaign.

- (f) *The authors say that "The ethical acceptability of control groups in situations perceived as threatening to patients (high endemic nosocomial infection and MRSA transmission rates) was an additional obstacle [to randomisation]". What arguments could be put against this view?* Having a control group would mean that a group of hospital staff would not experience the handwashing campaign. Their patients would not have the benefits of the campaign. The infections we are trying to prevent are potentially life-threatening. Thus if the campaign is effective in reducing infection, the patients attended by the control group staff will be at greater risk than those attended by treatment group staff. The counter arguments include the following. We do not actually know that the intervention will work, so it would be unethical to give it to everybody because if it did not work it would waste resources. The control group staff and hence their patients would receive what they would receive without the study, so would not be disadvantaged. Without a control group, it might be difficult to persuade others that the intervention is really effective, and so fewer staff and hence their patients might receive it than would be the case if the study were controlled. We are balancing the interests of the individual against the interest of the community. Here, as the authors say, randomisation of staff would be impracticable, but we could have control hospitals without the campaign but with observations. We could also randomise several hospitals to campaign or not.

Research and Critical Skills, 2002-3

Assessment 1

Read Course Notes Chapter 6 before attempting this assignment.

This is a formative exercise, designed to help you to learn and assess your own performance. Your answer should be typed. Please give your answer to your tutor by 14th November at the latest. Your answer will be marked by your tutor, who will give you feedback. You should receive this and a specimen answer before Assessment 2, which is summative for most students, is due.

Below is a recent research paper. Read the paper and answer the questions. This paper has not been edited and contains some statistical terms beyond the scope of the CFP course. You do not need to know anything about these statistical methods to understand the paper or to answer the questions. (Cardiac syndrome X was discovered at St. George's.)

The paper is by Raija Tyni-Lenne, Signe Stryjan, Björn Eriksson, Margareta Berglund Christer Sylven: Beneficial therapeutic effects of physical training and relaxation therapy in women with coronary syndrome X. *Physiotherapy Research International*, vol. 7, pages 35-43, 2002.

Answer the following questions. You should only need 2 to 6 sentences per answer.

1. The study is described as "single-blind" in the abstract, though no detail is given as to how this was done. What is meant by "single-blind"? What would the authors have to do to make this study single-blind? Why could the study not have been double-blind?
2. In Tables 2 and 3, means are given " \pm SD". What does "SD" mean here and what does it tell us about the data? Roughly how many observations will lie within the interval defined by mean \pm SD?
3. In Table 2, SDs are given for VO_2 . What features of the data suggest that there is a mistake in the SDs in the "before" column? (There are more than one.)
4. In Table 4, "median" and "range" are quoted. What do these terms mean? What do they suggest about the shape of the distribution of the Sickness Impact Profile score?
5. In Table 3, for distance walked by the physical training group compared to the non-training group a P value <0.003 is quoted. What does this mean? What can we conclude from it about whether physical training increases 6 minute walking distance in women with Coronary Syndrome X?
6. In Table 4, for the difference in Sense of Coherence score within the physical training group "NS" is quoted. What does this mean? What can we conclude from it about whether physical training increases Sense of Coherence score in women with Coronary Syndrome X?

Research and Critical Skills, 2002-3

Assessment 2

This is a summative assessment, which contributes to the total mark for this term. Although you may discuss the paper with your colleagues and tutor, the written answer must be your own work. You are reminded of the School's statement on plagiarism.

Your answer should be given to your tutor by 28th November at the latest. After that date, a specimen answer will be available on the Research and Critical Skills course web page.

Answers must be typed. Be sure to include your name and Research and Critical Skills group letter at the top of your answer. You should keep your own copy. This assignment will not be marked by your own tutor.

Read the following paper and answer the questions. This paper has not been edited and contains some statistical terms beyond the scope of the CFP course. You do not need to know these to understand the paper or answer the questions.

The paper is by Emmanuelle Girou, Sabrina Loyeau, Patrick Legrand, Françoise Oppein, Christian Brun-Buisson: Efficacy of handrubbing with alcohol based solution versus standard handwashing with antiseptic soap: randomised clinical trial. *British Medical Journal*, 2002, Volume 325, pages 362-366.

Answer the following questions. You should only need 2 to 6 sentences per answer.

- (a) "A total of 23 healthcare workers were included in the study and analysed; 12 were randomised to handrubbing and 11 to handwashing." What does "randomised" mean and why were the participants randomised to handrubbing or to handwashing?
- (b) "The microbiologist who examined the culture plates and reported the microbiological results was unaware of the hand hygiene method used." Why is this important?
- (c) In table 2, "Median (IQR)" for bacterial count in the handrubbing group before hand cleaning is "101 (29-380)". What is meant by the terms "median" and "IQR"? What can we deduce about the shape of the distribution of bacterial count, and why?
- (d) In the legend for figure 2, the authors report "difference between two groups significant (P=0.012)". What does this mean? What can we conclude about the effect of handrubbing and handwashing?
- (e) The authors say that for the reduction in bacterial contamination "The difference in the percentage reduction between the two groups was 26% (95% confidence interval 8% to 44%)". What is a confidence interval and what does this tell us about the effect of handrubbing and handwashing?
- (f) The authors conclude that "We have shown that handrubbing with an alcohol based solution is more effective than handwashing with an antiseptic soap in reducing bacterial contamination of healthcare workers' hands during routine patient care. This was due in part to the inadequate time spent washing hands conventionally." In what way do the results of the study support these conclusions?