1. Twenty patients with lung cancer were treated with a new form of radiotherapy whilst a further twenty were given standard radiotherapy. The survival times, measured in days from diagnosis, are given as follows:

New: 1, 33*, 47, 64*, 82, 82*, 116, 130, 150*, 164, 180*, 180*, 189*, 190*, 191, 192*, 192*, 198*, 198*, 198*

Standard: 17, 42, 44, 48, 60, 72*, 74, 95, 103, 108*, 122*, 144, 167, 170, 183, 185, 193, 195, 197*, 197*

* denotes a censored observation.

(a) Calculate the Kaplan-Meier estimate of the survivor function for each of the two treatment groups separately. Display the results graphically.

[10 MARKS]

(b) Test the null hypothesis that there is no difference between the treatments given to these patients, and interpret your results.

[7 MARKS]

(c) Suggest three different possible explanations for your findings.

[3 MARKS]

2(a) In a randomised trial of two treatments, where the primary outcome measure may be assumed to follow a Normal distribution, the number of patients required per group may be calculated from the formula $n = \left\{\Phi^{-1}\left(\frac{\alpha}{2}\right) + \Phi^{-1}(\beta)\right\} \frac{2\sigma^2}{d^2}$. Show how the formula may be derived, explaining what *n, I, ϑ, f*^{2} and *d* are.

[10 MARKS]

(b) A dietician is planning a trial of a new diet for serum cholesterol reduction. She thinks around 200 patients matching the clinical criteria can be found per year. Pilot studies suggest serum cholesterol has a standard deviation of 0.65mmol/l. To be clinically worthwhile, the new diet should reduce cholesterol by 0.2mmol/l when compared to a standard diet. Calculate the number of individuals that should be included to detect such a difference at the 5% level of significance with 90% power.

The dietician believes however that the new diet will reduce cholesterol by 0.4mmol/l when compared to the standard diet. Recalculate the number of individuals required for the same significance level and power as before.

Explain the difference between the results.

[5 MARKS]

(c) On the basis of your findings in (b), say how many patients you would advise to be included in the trial, justifying your answer.

State what further information you need before advising how long the study would need to recruit for and, by making a guess at this information, show how the length of the recruitment period could be calculated.

[5 MARKS]

3(a) The table shows the results from 14 trials comparing sclerotherapy to a control treatment for the reduction of mortality in patients with cirrhosis of the liver.

Trial	No of deaths / no patients	
	Sclerotherapy	Control
1	14/36	2/35
2	29/53	12/56
3	6/18	6/16
4	6/22	4/23
5	34/46	30/49
6	26/69	16/71
7	19/41	10/41
8	30/110	20/115
9	18/41	18/42
10	24/51	19/55
11	14/72	18/73
12	4/16	2/13
13	6/19	7/18
14	2/15	6/14

Explain what is meant by *publication bias.* Using graphical methods, investigate whether there is any evidence of publication bias in the example above.

[Note: var ($\log_e (OR_i)$) = $\frac{1}{a_i} + \frac{1}{b_i} + \frac{1}{c_i} + \frac{1}{d_i}$ where OR_i denotes the odds ratio in trial i,

a_i denotes the number of deaths on sclerotherapy in trial i,

b_i denotes the number of deaths on control in trial i,

c_i denotes the number of survivors on sclerotherapy in trial i, and

d_i denotes the number of survivors on control in trial i]

[13 MARKS]

Question 3 continued overleaf

Q3 contd

(b) The Mantel-Haenszel pooled odds ratio is given by $\hat{OR}_{MH} = \frac{\sum \frac{a_i d_i}{N_i}}{\sum \frac{b_i c_i}{N_i}}$ where a_i, b_i, c_i, d_i are as before and N_i denotes the total number of individuals in trial i.

Calculate $\hat{OR}_{MH}$ and comment on your findings.

[7 MARKS]

4. Given below are extracts from a published article by McGlone et al (1998), Journal of Accident and Emergency Medicine 15:231-236, 'An alternative to "brutacaine": a comparison of low dose intramuscular ketamine with intranasal midazolam in children before suturing.' The study is concerned with suturing (stitching) of lacerations (cuts) in children.

Give answers to the following questions using evidence from the article wherever appropriate.

(a) Define randomisation and state its purpose. Describe, in your own words, the allocation procedure adopted in this study and comment on its suitability for this sort of study.

[4 MARKS]

(b) State what is meant by 'blinding' in clinical trials. What form, if any, was used by the authors? Considering the outcome measures chosen, comment on whether potential biases in treatment comparisons could have been lessened in this study.

[4 MARKS]

(c) Describe what is meant by an 'intention-to-treat' analysis. How many children were randomised in this study? How many were included in the analysis of the amount of restraint needed? Comment on the potential for bias in this analysis.

[2 MARKS]

(d) State what is meant by 'external validity'. Using specific evidence from this paper, comment on the external validity of this study.

[3 MARKS]

(e) Does the imbalance in the percentage of co-operative children in each group pre-allocation (Table 2) affect the conclusions of the study? Justify your answer.

[2 MARKS]

(f) Explain what is meant by 'non-response bias'. Comment on the potential for non-response bias in this study.

[2 MARKS]

(g) Comment on the overall quality of the study design and the validity of the authors' conclusions justifying your answer. [3 MARKS]

5(a) Define the terms *sensitivity*, *specificity*, *positive predictive value* and *prevalence of disease*, as used in the context of diagnostic testing.

Using Bayes' theorem, show how positive predictive value is related to the other three quantities.

[8 MARKS]

(b) The following data are concerned with the use of temperature as a screening test for acute appendicitis in patients admitted to a hospital casualty department with abdominal pain.

	Appendicitis		
Temperature	Yes	No	Total
>37.5°C	44	58	102
♋37.5°C	19	89	108
Total	63	147	210

(i) Calculate the sensitivity and specificity of 'temperature greater than 37.5°C' as a marker for appendicitis.

(ii) Calculate the positive predictive value assuming the patients to be a random sample from the population to be screened.

(iii) Recalculate the positive predictive value if the prevalence in the population to be screened is 20%.

Comment on your findings.

[7 MARKS]

(c) Calculate a 95% confidence interval for the sensitivity. If a clinician decides this study is not sufficiently precise, how many individuals with appendicitis would be needed in order to estimate the confidence interval with an absolute margin of error of 5% either side?

[5 MARKS]

6(a) Describe the main features of (i) a cohort study and (ii) a case-control study. Give two advantages and two disadvantages of each design when used to examine the association between a risk factor and some outcome.

[6 MARKS]

(b) Cystic fibrosis is a condition affecting mainly the digestive system and the lungs. It is usually diagnosed soon after birth and symptoms occur throughout life. A paediatrician caring for children with cystic fibrosis noticed a recent increase in the incidence of fibrosing colonopathy in her patients. Fibrosing colonopathy is a condition affecting the colon which can cause abdominal pain. It is usually treated by surgery but fortunately is a rare problem. The paediatrician hypothesised that the increase in incidence was related to the coating of a pill, mexonate, recently introduced onto the market as part of the treatment for cystic fibrosis. Mexonate can be given in varying doses but is not necessarily given to all children.

Outline how you would design

(i) a cohort study

(ii) a case-control study

to examine the paediatrician's hypothesis.

[8 MARKS]

(c) Which of these study designs would you use to test this hypothesis and why?

[2 MARKS]

(d) Bradford-Hill set out the following nine criteria to be used in assessing the evidence for causality: strength of association, consistency of evidence, specificity, temporal relationship, dose-response relationship, biological plausibility, coherence of evidence, experimental evidence, analogy. Choose four of the criteria and explain how they might be used in the above example to assess the evidence that mexonate causes fibrosing colonopathy in cystic fibrosis children.

[4 MARKS]

7(a) Describe Bayesian inference giving definitions of the terms *likelihood*, *prior distribution* and *posterior distribution*. With reference to a Bayesian analysis of a clinical trial comparing two treatments, describe possible sources of prior information.

[6 MARKS]

(b) In a clinical trial comparing sclerotherapy with a control treatment with respect to one-year mortality it is assumed that the death rate on the control group, Π_c, follows a Beta distribution with parameters I_c and ϑ_c. Under this assumption the mean and variance of Π_c are then given by

$$E(\theta_c) = \frac{\alpha_c}{\alpha_c + \beta_c} \qquad \text{and} \qquad V(\theta_c) = \frac{\alpha_c \beta_c}{(\alpha_c + \beta_c)^2 (\alpha_c + \beta_c + 1)}$$

respectively.

Find expressions for I_c and ϑ_c in terms of $E(\Pi_c)$ and $V(\Pi_c)$.

[2 MARKS]

(c) Summarising previous evidence suggests that for the control treatment, one-year mortality is around 20% with approximately 95% certainty that it is no less than 16% and no more than 24%. Sclerotherapy is believed to reduce mortality to around 15%, with corresponding 95% certainty that it is no less than 7% and no more than 23%.

Express your current beliefs about the log of the odds ratio for one-year mortality of sclerotherapy compared to control treatment and calculate a 95% credibility interval for this quantity.

[Hint: If $\Pi \sim$ Beta (I, ϑ), $\log(\Pi/(1-\Pi)$ is approximately $N\left(\log\left(\frac{\alpha - 0.5}{\beta - 0.5}\right), \quad \frac{1}{\alpha} + \frac{1}{\beta}\right)$]

[5 MARKS]

Question 7 continued overleaf

Q7 contd

(d) The trial was conducted and the results were reported as follows: of 164 patients randomised to sclerotherapy 40 had died by one year whilst on control 32 out of 158 randomised patients had died. Estimate the distribution of the log of the posterior odds ratio and calculate a 95% credibility interval using the prior information described above. Comment on the posterior compared to the prior beliefs.

[Hint: if there are r_c deaths out of n_c patients on control, and r_s deaths out of n_s on sclerotherapy, the log of the posterior odds ratio is approximately Normally distributed with mean T and variance f^2 given by

$$\mu = \log\left[\frac{(\alpha_s + r_s - 0.5)(\beta_c + n_c - r_c - 0.5)}{(\alpha_c + r_c - 0.5)(\beta_s + n_s - r_s - 0.5)}\right]$$

$$\sigma^2 = \frac{1}{\alpha_c + r_c} + \frac{1}{\beta_c + n_c - r_c} + \frac{1}{\alpha_s + r_s} + \frac{1}{\beta_s + n_s - r_s} \quad]$$

[7 MARKS]