Written Examination for Design and Analysis of Clinical Trials (MSA620)

Date and place: 2016-03-17,
Examiner: Ziad Taib (0707655471)
Rules: This is a closed book exam. No material is allowed other than a simple pocket calculator and statistical tables.
Scores: The written exam is worth 30 scores. There are three possible grades: Excellent (VG) (at least 24 scores), pass (G) (at least 15 cores) and do not pass (U) (less than 15 scores). In addition to the written exam, the computer assignments are compulsory.

---

1. Six different types of bias might skew or corrupt the conclusions from data.
    a. Define at least four of these types. (2p)
    b. Choose three of these and explain what consequences they might have in the context of clinical trials. (2p)

2. Crossover is a common type of design.
    a. Explain what it is by giving an example (including a picture). What are the pros and cons with a crossover design, and when is it suitable/unsuitable to use? (2p)
    b. What is a balanced crossover design? Give examples of two different types of balanced designs. Assume you want to compare four treatments in a balanced design. Set up one design of each kind (draw graphs). (2p)

3. To determine the sample size of a clinical trial you need to consider a number of factors. Mention at least four such factor and explain what they stand for. (2p)

4. Describe what kind of mistake you do if you make a type 1 and type 2 error, respectively and explain the meaning of "family wise error rate". (2p)

5. In a clinical trial, males with advanced inoperable lung cancer were randomized to a standard therapy and a test chemotherapy. The primary endpoint for the therapy comparison was time to death in days, represented by the variable *Time*. The data include information about a number of explanatory variables: *Therapy* (type of therapy: standard or test), *Cell* (type of tumor cell: adeno, large, small, or squamous), *Age* (age, in years), and *Duration* (months from diagnosis to randomization). The variables *Cell*, and *Therapy*, which are categorical variables, are declared in the CLASS statement. Here *Cell*=large is chosen as the reference category for type of tumor cell, and *Therapy*=standard is chosen as the reference category for the type of therapy. The following SAS code was used to analyse these data and the results from the analysis are shown in Table 3 below.

```
proc phreg data=VALung;
class Cell(ref='large') Therapy(ref='standard');
model Time*Status(0) = Duration Age Cell Therapy;
run;
```

    a. Which of the cell types has a significant association with increased risk of death? (1p)
    b. Calculate the missing hazard ratio for Therapy and explain how it should be interpreted. (2p)

6. We are planning a dose finding clinical trial involving three doses of a new drug (Placebo, a low dose D1 and a high dose D2) and two endpoints E1 and E2. There is a proposal to compare D1 with placebo with respect to E1 and in case that comparison turns out to be significant proceed to make the comparison in terms of E2. If also the latter comparison is significant, we do the same thing for D2.
    a. How should we perform these tests in order to maintain an overall alpha level of 0.05? Do you see any problems with this dose testing strategy? (draw a graph) (2p)
    b. Can you think of an alternative strategy? What good features does your new strategy have? (2p)

7. In a clinical trial of a respiratory disease, patients are randomized into one of two treatment (Drug and Placebo) in each of two centres. During the treatment period, respiratory status is determined at four different visits as a binary variable (zero=poor and one=good). 54 thus received the active drug and 57 were given placebo.
    a. As a first simple analysis, the mean responses over time in the two arms are compared using a t-test. What do you think of this approach? (1p)
    b. A second approach uses the endpoint: the number of visits at which the patient had status 1, assuming this endpoint follows the binomial distribution on the individual level. Do you think this is an appropriate endpoint? (1p)
    c. To be able to account for some covariates (sex, age, baseline value, centre), a model with logistic link function was proposed for the response at visit 4. Results from this analysis are shown in Table 1.
        i. What conclusions do you draw from these results? (2p)
        ii. Find a 95% confidence interval of the odds ratio of the treatment effect. What does that CI say about the effect of the drug? (1p)
    d. The above model does not use all the responses. Therefore, we use a unified logistic model for the response of the ith subject at the jth visit. The output from that model is in Table 2.
        i. How does the unified model look like? (1p)
        ii. What conclusion do you draw from these values? Do you see any problem with this model? (2p)

8. What does center-treatment interaction mean? Assume a statistical model for response data from a multi-centre trial including factors for grand mean center, treatment, center-treatment interaction. Which factors should (according to you) be (i) fixed and (ii) random? Motivate your choices. (3p)

Tables:

Table 1. Result from GLM for a logit model			
Covariate	**Estimate Odds Ratio**	**SE**	**p-value**
Treatment	3.544	1.208	<0.001
Baseline	6.333	2.197	<0.001
Age/100	0.153	0.196	0.1
Sex	1.147	0.488	0.7
Centre	1.915	0.661	0.06

Table 2. Parameter estimates from a logistic regression model		
Covariate	**Estimate Regess Coeff.**	**SE**
Treatment	1.267	0.235
Time	-0.078	0.099
Sex	0.137	0.294
Age/100	-0.189	0.883
Baseline	1.849	0.240
Centre	0.651	0.238

Table 3. Parameter estimates from a Cox regression model							
		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
Duration		1	0.0032	0.0095	0.12	0.7335	1.003
Age		1	-0.0135	0.0096	1.98	0.1597	0.987
Cell	**adeno**	1	0.7836	0.3038	6.65	0.0099	2.189
Cell	**small**	1	0.4823	0.2654	3.30	0.0691	1.620
Cell	**squamous**	1	-0.4077	0.2836	2.07	0.1506	0.665
Therapy	**test**	1	0.5666	0.2476	5.23	0.0221	XX