MAS6061 Project 2 Epidemiology 2017/18

Students should send answers in Word or PDF documents. This assignment should be submitted via MOLE on Tuesday 30th January 2018 by 12:00. **These exercises form part of the assessment of the module (Total marks of 90, worth 70% of the mark for Epidemiology).**

**To obtain high marks in this project students need to justify which methods they have selected to use in each case and offer appropriate discursive interpretations of results.**

1. For patients with stable heart disease, visual assessment of a coronary angiogram is a poor estimator of the physiological significance of coronary artery lesions. Fractional flow reserve (FFR) is an invasive medical procedure which measures the degree of obstruction to flow in conditions of hyperaemia. This measurement is reported as a proportion and its values must lie between 0 and 1. However, in clinical practice more than 95% of patients do not have an FFR measurement (mFFR). So research teams have developed computer models which estimate FFR (virtual, or vFFR) from angiographic images. One model (SISA) has been used on a consecutive series of 101 patients where the vFFR was also measured. (Project2.xls sheet 1)

You are asked to analyse these data for agreement and clinical utility.

a) What is your assessment of the vFFR SISA measurement as a possible alternative to the presumed best standard mFFR? **[10 marks]**

b) In reality the mFFR is used to make clinical decisions. In prior clinical evaluations a mFFR above 0.75 would make a patient likely to benefit from a STENT operation. Hence the critical threshold of 0.75 is often used in practice. In the light of this new information examine the impact (positives and negatives) of using the vFFR SISA where the mFFR is not available to guide clinical decisions to performing a STENT operation. **[10 marks]**

2. Project2.xls (sheet 2) contains data on 2726 patients who were admitted to an emergency department (ED) as part of a study of head injuries where the patient was taking an anticoagulant. All patients were followed for up to 6 weeks from admission and any poor outcomes were recorded (e.g. death or intracranial bleed). For each patient the following data was recorded; their age, gender, INR value at admission, Glasgow Coma Scale at admission, whether the patients reported any of the four symptoms (amnesia, loss of consciousness, vomiting, headache) following the head injury. The data in the spreadsheet shows patient information:

Age, gender(1=male), INR, GCS, amnesia, vomiting, loss-conscious, headache, CT-performed, OutcomeCode.

INR is a normalised ratio measure of how long it takes for blood to clot. Values between 2-3 are common for patients who are taking anticoagulants. The higher the INR is, the longer time it will take for a clot to form and hence this could have very serious consequences for bleeding following a head injury. The INR value indicates how ‘thin’ the blood is and hence a high value would be believed to put the patient at a raised risk of prolonged bleeding. The Glasgow Coma Scale is a simple scoring system which summarises how responsive and conscious a person is. A score of 15 is the highest level of consciousness and 3 is the lowest level meaning the patient is completely unresponsive. Many patients were offered CT scans soon after arrival at the ED. An early CT scan is recommended for patients when GCS was below 14 and if vomiting had been reported. (The CT scan cannot prevent the poor outcome but it may avoid serious consequences associated with a poor outcome (e.g. death))

a) What factors (variables) seem to influence which patients are offered a CT scan on arrival? Is the data consistent with hospitals adhering to the recommendation using CGS<14 and a report of vomiting? **[8 marks]**

b) Which variables are associated with a poor outcome within 6 weeks? For the continuous/numerical variables is it fair to assume they show a linear relationship with risk of the poor outcome? **[10 marks]**

c) Build a multivariable predictive model (using the Likelihood Ratio Test to select optimal model with alpha=5%) to predict a poor outcome for patients presenting at an emergency department following a head injury when taking anticoagulants. In this application we are keen to find an optimal clinical decision tool which will identify those patients who will not likely benefit from a CT scan. Hence we would weight sensitivity more valuable than specificity in this context. Explore the calibration of your ‘best’ model and report the sensitivity and specificity. **[14 marks]**

d) Another research group has developed a predictive model for a similar situation but limited to predict risk of a poor outcome where GCS score is high (i.e. 15). Their equation for the logit(risk)=0.83 - 0.82 x headache -0.67 x amnesia. Their result is somewhat controversial as they do not include ‘vomiting’ as a predictor in this model. How does this predictor perform (calibration and discrimination) on the relevant subset of the study in sheet 2? **[8 marks].**

3. Sheet 3 contains data collected on a cross sectional study of *1879* Rheumatoid Arthritis (RA) patients. A panel of 7 genetic markers (SNPs) in candidate genes have been selected for investigation in this series following a genomewide association study of susceptibility to rheumatoid arthritis. These loci were all associated with susceptibility to RA so this research team wants to know if the same loci are associated with severity of disease. Prior studies have demonstrated that severity of arthritis (debilitating accumulation of erosions) is associated with smoking, age and duration of disease. If good genetic predictors of increased risk of severity can be identified this may mean patients with specific high risk genotypes can be offered more aggressive treatments.

a) Perform standard quality control checking on the genotyping of the 7 candidate genes measured in this cross sectional study. State all of your assumptions and declare which (if any) patient samples or SNPs need to be discarded for quality control reasons**. [10 marks]**

b) Examine the association between the SNPs and for their association with RA severity. Adjust the analysis for the known risk factors, age, duration of disease and smoking. Calculate the population attributable fraction/risk (based on the equation 3.23 on page 112, Woodward textbook) for each SNP that is statistically significantly associated with severity. State any assumptions you have made in your analysis and comment on your results. **[10 marks].**

c) Could any of your conclusions be affected by the presence of a confounding or modifying variable? Explore if this could be the case in this cross sectional study. Discuss your results and modify your conclusions to b) if necessary. **[10 marks]**