The impact of existing comorbidities on the mortality of hospitalized Covid-19 patients

This exam is open-book but must be completed independently. You may not discuss the problems with anyone other than the instructor (lmao@biostat.wisc.edu).

Clinical and epidemiological evidence suggests that covid patients endure worse outcomes if they have pre-existing conditions. To study the specific impact of various comorbidities, consider a cohort of 582 covid patients admitted between January and August 2020 in two hospitals in the UK. The study data contain demographic and clinical variables extracted from the hospital electronic records, along with the patient's survival status in the first 30 days after diagnosis. The **primary objective** is to identify and quantify the effects of the most impactful comorbidities (e.g., metabolic, cardiovascular, respiratory, cardiovascular, renal, or cancer) on the mortality of hospitalized covid patients. A secondary objective is to assess other socio-demographic risk factors for covid mortality, such as age, ethnicity, and/or socioeconomic status.

The dataset of interest is in `Files//Midterm//combd_covid19.csv`, with the following key variables:

`id:` Patient identifier
`time:` Time (days) from diagnosis (hospital admission) to death or censoring
`death:` Yes = death in the first 30 days; No = discharge/transfer/loss to follow-up
`sex:` Sex (Female and Male)
`agecat:` Age group (years)
`ethnicity:` Ethnicity (White, South Asian, and Other)
`imd:` The English Indices of Multiple Deprivation (IMD; 1—5: most to least deprived)
`bmicat:` BMI category (Healthy weight, Overweight, and Obese)
`==========` Variables for comorbidities `==========`
`diabetes1:` Type I diabetes (Yes v No)
`diabetes2:` Type II diabetes (Yes v No)
`hypertension:` Hypertension (Yes v No)
`cvd:` Cardiovascular disease (Yes v No)
`asthma:` Asthma (Yes v No)
`copd:` Chronic obstructive pulmonary disease (COPD; Yes v No)
`cancer:` Cancer (Yes v No)
`renal:` Renal disease (Yes v No)
`==========`

1. **Descriptive statistics:**
   Summarize the demographic (sex, age, ethnicity), socioeconomic (IMD), and clinical (BMI and pre-existing comorbidities) variables **by sex and overall**. Be sure also to calculate the death rates.

2. **Graphical analysis**:
   As preliminary analysis, compute and plot the Kaplan—Meier survival curves marginally by sex, age group, ethnicity, IMD, BMI categories, and status of each of the comorbidities. For the comorbidities, if appropriate, combine type I and II diabetes into one variable for diabetic disease (yes v no), hypertension and CVD into one for cardiovascular disease (yes v no), asthma and COPD into one for pulmonary disease (yes v no).

3. **Multiple regression analysis:**

   Build a Cox proportional hazards regression model to analyze the effects of existing comorbidities on patient survival adjusting for age, ethnicity, and IMD. Some specific guidelines:
   a. Based on the graphical results produced in step 2, pick relevant covariates as appropriate;
   b. Check the proportionality assumption on the each of the covariates included and take measures to address possible violations;
   c. Report the analysis results under your final model. Which comorbidities are most strongly associated with Covid mortality and to what extent? In addition, conduct formal tests on the effects of age and IMD scores (two separate chi-square tests, each with multiple degrees of freedom).
   d. Fit another model with the total number of comorbidities as the main predictor in place of the individual comorbidities (still adjusting for socio-demographic variables). Explore the appropriateness of entering the predictor in linear form. Report its effect on patient mortality.

4. **Reporting:** Write a short paragraph (3—4 sentences) summarizing the most important findings of the analysis in the form of a press release for the mass media. Be certain to use concise, nontechnical language that would be understandable to a lay reader, while still accurate scientifically.