DATA MANAGEMENT INTERVIEW.

Dataset name: bl_demog.dta
Description: Stata dataset containing patient id, baseline demographic information plus some lifestyle, vital signs, anthropometric and laboratory measurements.

Variable name	Variable description	Coding etc.
ptid	Unique patient identifier	String length 9
birthdt	Date of birth	Stata date format
age	Age (years)	Numeric
agegroup	Age categories	0=30-64, 1=65-69, 2=70-74, 3=75+
sex	Sex	String: Female, Male
smkstat	Smoking status (5 levels)	1=Never, 2=Ex-Light, 3=Ex-Heavy, 4=Current-Light, 5=Current-Heavy
smoke	Smoking status (3 levels)	String: Current, Ex, Never
race	Ethnic group	String: Asian, Black, White, Other
hfdiag	Heart failure diagnosis	1=Ischemic, 2=Non-ischemic
wt	Weight (kg)	Numeric
ht	Height (cm)	Numeric
wc	Waist circumference	Numeric
wc_unit	Waist circumference	String: CM, M
sbp	Systolic Blood Pressure (mmHg)	Numeric
dbp	Diastolic Blood Pressure (mmHg)	Numeric
hrate	Heart Rate (bpm)	Numeric
egfr	Estimated Glomerular Filtration Rate (ml/min/1.73m)	Numeric
lvef	Left ventricular ejection fraction (%)	Numeric
diab	Diabetes	0=No, 1=Yes

Section 1:

1. How many participants were enrolled in this study?

2. How many continuous and categorical variables are in this dataset?

3. What was the average age for males and females in the study?

4. What is the distribution of; age, height, weight, and Estimated Glomerular Filtration Rate?

5. Was there an independent association between smoking, age, egfr, diabetes, sex with heart failure?

6. Where there any factors that confounded the relationship between heart failure and smoking?

Section 2

7. In 1854, John Snow compared cholera deaths across areas supplied by different water providers (1). His findings are below.

Local water provider	Population	Cholera deaths from 8 July to 26 Aug 1854	Incidence risk per 100,000
Southwark and Vauxhall Company	167,654	844	
Lambeth Company	19,133	18	
Total			

a. What are the comparison groups for John Snow's investigation? State the null hypothesis.

b. Complete the bottom row and calculate an overall incidence risk of cholera deaths.

c. Calculate the incidence risk of cholera deaths for each local water provider. Report this as deaths per 100,000 populations.

8. A study was set up to investigate trends in early childhood death across Europe. The researchers examined national all-cause mortality rates, and cause-specific mortality rates, for children aged under 5 years. They plotted mortality rates by calendar period for each country.
   a. Study design?
   b. Study population?
   c. Primary outcome?
   d. Primary exposure?
   e. Measure of disease (outcome) occurrence?
   f. Possible measures of exposure effect?