GLM 1: Practical session

In this practical session you will be given a data set and asked to fit and interpret a series of GLMs.

HINT: use

```
glm <outcome> <explanatory vars>, family() link() eform
```

to return exponentiated parameters and their 95% C.I.s in the model output.

The data can be found on the **Advanced Statistics for Records Research** moodle page and are in asrr_glm_1.dta. The file contains simulated (i.e. made up) data for 5000 people covering a study period of 1 year. This is all data that could be derived from administrative hospital records.

Variable Name	**Variable Label**	**Variable Description**
id	person id	unique person identifier from 1 to 5000
age	age (year)	age in years at beginning of the study period
sex	sex	1 = Female 0 = Male
cvd	History of Cardiovascular Disease	1 = diagnosis of cardiovascular disease in the year before the study period 0 = no diagnosis
imd5	IMD quintiles	Index of multiple deprivation (IMD) quintiles derived from patient addresses (postcodes). 1 = most deprived, 5 = least deprived
ae_atd	A&E attendances	number of A&E attendances over the study period
death	Death	1 = died over the study period 0 = alive at the end of the study period

1. Using the `tabulate <var>` (or `tab`) or `summarize <var>, detail` (or `su <var>, d`) command in stata, complete the table below to get a better idea of what is in the data file (e.g. `tab imd5 or su age, d`) :

Patient characteristic	**Overall (n = 5000)**
Age (years), median (min, max)	
Sex (1 = Female) (%)	2295 (45.9)
IMD quintiles (%)	
1	962 (19.2)
2	1019 (20.4)
3	
4	1040 (20.8)
5	
History of Cardiovascular Disease (%)	
Deaths (%)	2111 (42.2)
A&E attendances (median [range])	2 (0, 11)

2. Now we're going to look at how the risk of A&E attendance over the study period varies by age and sex:
    a. Using `graph bar, over(ae_atd)`, look at the distribution of number A&E attendances over the whole sample.

    b. How does its distribution differ by sex? (HINT: add the option `by(<var>)` to your stata command)

    c. Use an appropriate GLM to confirm whether the number of A&E attendances differs by sex (remember to add the prefix `i.` when including binary or factor variables in your modelling)

Sex	Rate Ratio	95% C.I.
Female		

    d. Add age to the model in (c). Interpret the association between number of A&E attendances and age, adjusting for sex:

e. Finally, add cvd to the model in (d). Interpret the association between number of A&E attendances and History of Cardiovascular disease, adjusting for age and sex:

3. Now we will look at how the likelihood of death varies over the study period by some of the other patient characteristics.

   a. How many people died, by IMD quintile?
   `tab imd5 death, row`

IMD quintile	Number that died	%
1		
2		
3		
4		
5		

   b. Using a GLM, model the number of deaths by IMD quintile and completed the table below:

IMD quintile (compared to 1)	Odds Ratio	95% C.I.
2		
3		
4		
5		

c. Interpret these findings (i.e. the parameter estimates and their 95% C.I.s)?

d. Add in age and sex to the model. Describe what happens to the parameter estimates for imd5 (in one sentence)?

e. Add cvd into the model in (e). Describe what happens to the parameter estimates for imd5 now (in one sentence)?

Practical end.

Solutions and code (stata and R) will be available on the course moodle page at the end of the practical.