Multi-variable regression

Nuredin Mohammed

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Based on slides Camille Maringe

Recap

- Exposures / Outcomes
- Types of variables
    - **Quantitative**: birth weight; height; haemoglobin level …
    - **Categorical**: age groups; smoking status; treatment/placebo …
- Hypothesis testing
    - Quantitative variables (comparing means): t-test
    - Categorical variables (comparing proportions): Chi-square test
- Measures of effect
    - Odds Ratio; Risk Ratio
    - Rate ratio
    - Risk difference

Confounding?

Effect modification?

Stratified measures of effects

Recap

- Modelling
    - Quantitative outcome: linear regression
    - Binary outcomes: logistic regression
    - Time-to-event outcomes (binary status + follow-up time): Cox, Poisson regression

Confounding? Multi-variable regression

Effect modification? Multi-variable regression with interaction terms

Recap: (univariable) Cox/Poisson regression

- Cohort data with follow-up information

	Incident cases of death	Person-years at risk
Normal vision	97	10,625
Visually impaired	40	832

- Rate Ratio:
  - Rate of death in visually impaired: 40/832=0.0481 = 48.1 deaths /1,000 person-years
  - Rate of death in normal vision: 97/10,625=0.009129 = 9.1 deaths /1,000 person-years
  - Incidence Rate Ratio of visual impairment: 48.1/9.1 = 5.28

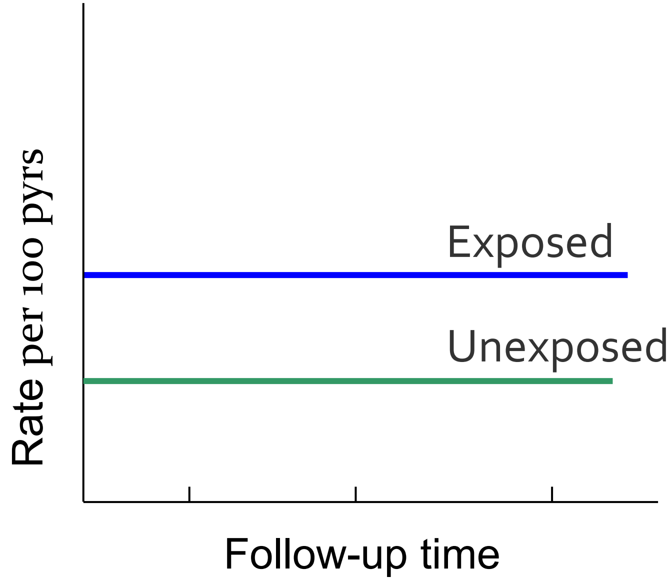
Rates Always positive

Transform: rates >> log(rates)

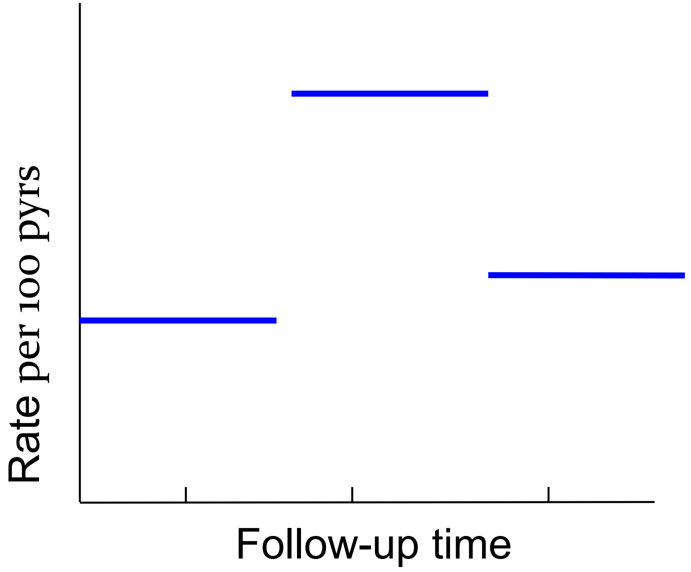
$$\log(\mathit{rates\ in\ exposed}) = \log(\mathit{rates\ in\ unexposed}) + \boldsymbol{log}(\boldsymbol{RR})$$

Recap: (univariable) Cox/Poisson regression

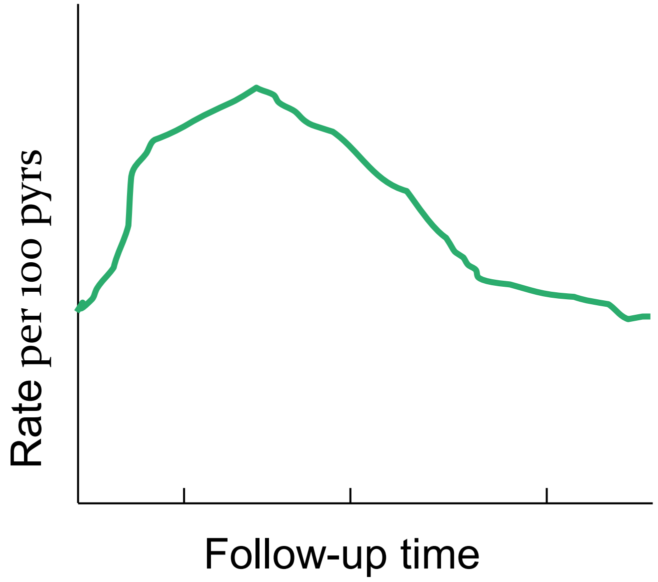
Poisson regression:
Rates are constant through time



Poisson regression:
Split/cut follow-up time



Cox regression:
Non-parametric baseline rate (unexposed)
Rate Ratio constant



Objectives

- Understand how the univariable regression can be extended to two or more explanatory variables
- Understand the uses of multi-variable regression
- Know how to conduct and interpret a multi-variable regression

Multi-variable regression is an extension of univariable regression which allows us to assess the effect of more than one explanatory variable on the response variable simultaneously.

From univariable regression…

$$\hat{y} = \alpha + \beta x$$

	Linear regression	Logistic regression	Cox/Poisson
$\hat{y}$	Outcome	Log Odds(outcome) in exposed	Log Rate(outcome) in exposed
x	Value of the exposure	Value of the exposure	Value of the exposure
α	Mean outcome in unexposed (or exposed = 0)	Log Odds(outcome) in unexposed	Log Rate(outcome) in unexposed
β	Effect of exposure on outcome	Log Odds Ratio	Log Rate ratio

To multivariable regression

$$\hat{y} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots\ \beta_k x_k$$

	Linear regression	Logistic regression	Cox/Poisson
$\hat{y}$	Outcome	Log Odds(outcome) in exposed	Log Rate(outcome) in exposed
$x_1, \dots x_k$	Values of explanatory variables	Values of explanatory variables	Values of explanatory variables
α	Mean outcome for all explanatory variable set to 0	Log Odds(outcome) when all explanatory variable set to 0	Log Rate(outcome) when all explanatory variable set to 0
β_1	Effect of exposure x_1 when all other variables are held constant	Log Odds Ratio for x_1 when all other variables are held constant	Log Rate Ratio for x_1 when all other variables are held constant
β_k	Effect of exposure x_k when all other variables are held constant	Log Odds Ratio for x_k when all other variables are held constant	Log Rate Ratio for x_k when all other variables are held constant

Poisson regression: visual impairment and death

- **Example:** Is visual impairment associated with varying rates of death after adjusting for confounding effect of microfilarial infection?

- Cohort study conducted in Nigeria (bab.dta).
- Visual impairment: normal vision 0/ impairment 1
- Microfilarial infection: categorised in 4 levels (negative; <10; 10-49, ≥50 mf/mg)
- Outcome is both (i) vital status (137 individuals die) & (ii) time to death/censoring

Multivariable Poisson regression

>> effect of vimp

```
prm_died_vimp <- glm(died ~ vimp + offset(log_p_years),
                     family = poisson(),
                     data = mortality)

coeftest(prm_died_vimp)
```

Likelihood Ratio Test in R

prm_died_vimp >> **model with only effect of vimp**

	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Visually impaired	**5.263477**	.9890475	8.84	0.000	3.641879	7.607115
_cons	.0091295	.000927	-46.25	0.000	.007482	.0111397

Multivariable Poisson regression

```
>> effect of vimp and agebin

prm_died_vimp_agebin <- glm(died ~ vimp + agebin + offset(log_p_years),
                            family = poisson(),
                            data = mortality)

coeftest(prm_died_vimp)
```

Multivariable Poisson regression

prm_died_vimp_fup >> effect of vimp and follow-up

	Haz. ratio	Std. err.	z	P>|z|	[95% conf. interval]
vimp					
Normal	1				
Visually ..	2.597216				
agebin					
15-54	1				
55+	4.018964				
_cons	.0073068				

RR for the effect of vimp, adjusted for age : 2.597

RR for the effect of age (55+ *vs.* <55), adj. for vimp : 4.019

Rate in unexposed (vimp = 0 and age = 15-54): 0.0073

Likelihood Ratio Test

Is adjustment for *age* adding to our understanding of rates of death?

```
> lrtest(prm_died_vimp, prm_died_vimp_agegrp)
Likelihood ratio test

Model 1: died ~ vimp + offset(log_p_years)
Model 2: died ~ vimp + agebin + offset(log_p_years)
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   2 -691.49
2   3 -669.58  1 43.831   3.58e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Some final thoughts: assumptions and limitations

- Regression allows many explanatory variables to be included
  - !!! beware of spurious associations

- Plan the analysis:
  1. What are your hypotheses
  2. Which explanatory variables
  3. Which interactions

  !!! Prioritise

- Check for non-linearity by defining categorical variables.

Recap

Outcome	Modelling Approach
Continuous	Linear regression
Binary	Logistic regression
Rate/Time to event	Poisson regression/Cox-regression

Thank you!

LONDON
SCHOOL of
HYGIENE
&TROPICAL
MEDICINE