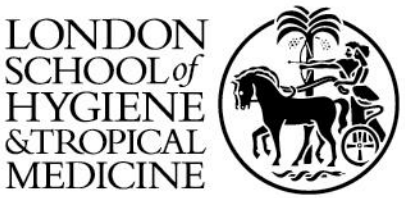
Introduction to Poisson and Cox regressions

Nuredin Mohammed



Based on slides by Craig Higgins

Learning Objectives

At the end of this lecture, participants should be able to:

- Estimate a crude rate ratio
- Carry out significance tests (Wald, LRT)
- Explore a dose response for an ordered categorical variable

Also

- Appreciate when a Cox or Poisson model is preferable

Where does this lecture fit? (1/2)

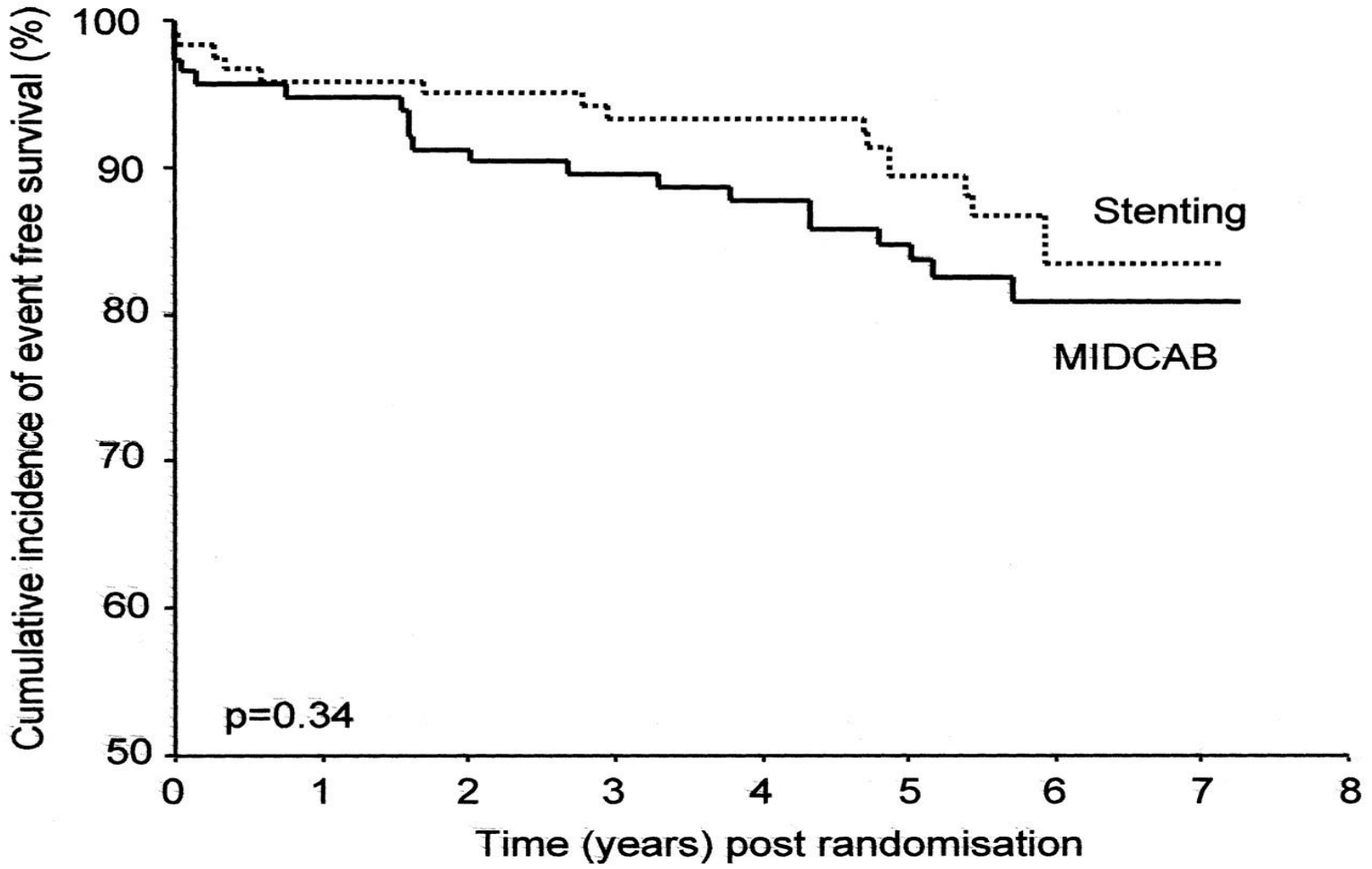


Figure 4. Kaplan-Meier curve showing freedom from death and myocardial infarction. Patients assigned to stenting are indicated by a dotted line and those assigned to minimally invasive direct coronary artery bypass surgery (MIDCAB) by a solid line.

Where does this lecture fit? (2/2)

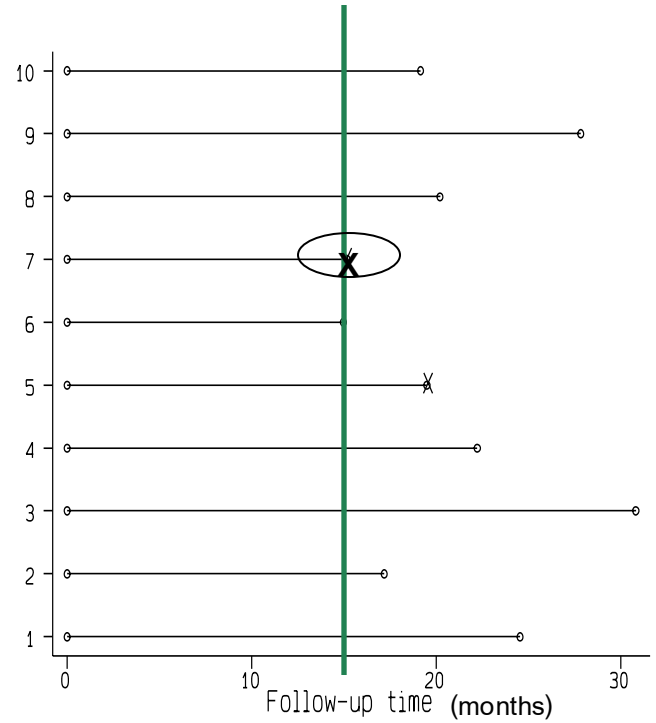
Parametric modelling for different types of outcomes:

Linear regression models	<<	continuous outcomes
Logistic regression models	<<	binary (or categorical) outcomes
Poisson	<<	time-to-event outcomes
Cox	<<	time-to-event outcomes

Example

- Cohort study conducted in Nigeria
  - Visual impairment measured at recruitment
  - Participants followed-up for ~ 3yrs
  - Outcome: death
- Date of entry and exit from cohort

id	agegrp	vimp	enter	exit
1	55-64	Normal	09may1989	05feb1992
2	35-54	Normal	09may1989	05feb1992
15	55-64	Visually impaired	11may1989	05feb1992
38	15-34	Normal	15may1989	05sep1991
319	15-34	Normal	13may1989	06sep1991



Estimation of Rates and Rate Ratios

	Number of deaths	Person-years at risk
Normal	97	10,625
Visually impaired	40	832

Mortality rate 'normal' group $= \frac{97}{10{,}625} = 0.0091 = 9.1\ deaths\ per\ 1{,}000\ pys$

Mortality rate 'visually impaired' group $= \frac{40}{832} = 0.0481 = 48.1\ deaths\ per\ 1{,}000\ pys$

Mortality rate ratio $= \frac{48.1}{9.1} = 5.26$

Estimation of Rates and Rate Ratios

To calculate rates (per 1000) in each group:

```
inc_vimp <- mortality %>%
  group_by(vimp) %>%
  summarise(
    D = sum(died),
    Y = as.numeric(sum(exit - enter)/365.25),
    rate = D / Y, Rate = rate * 1000
    )
```

Estimation of Rates and Rate Ratios

Rates:

vimp	D	Y	Rate
Normal	97	10.6249	9.1295
Visually impaired	40	0.8324	48.0527

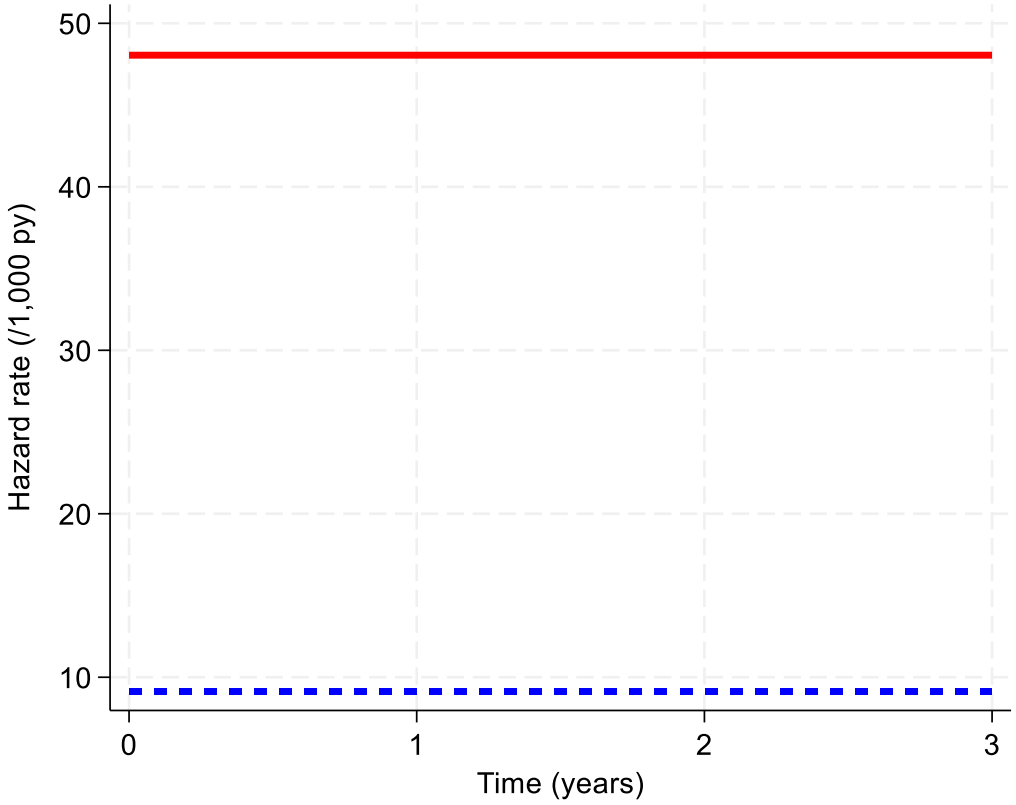
Estimation of Rates and Rate Ratios

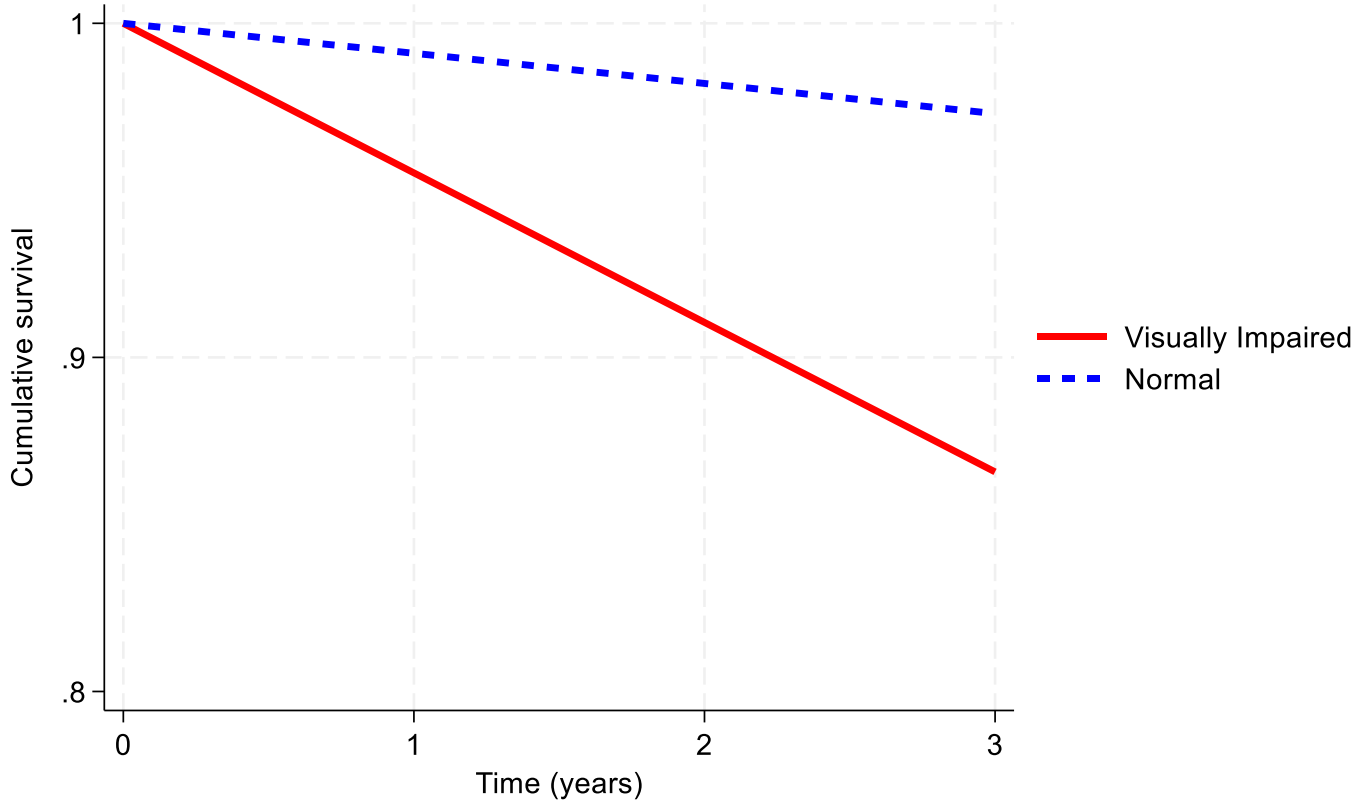
Rates:

vimp	D	Y	Rate
Normal	**97**	**10.6249**	**9.1295**
Visually impaired	**40**	**0.8324**	**48.0527**

RR	chi2	P>chi2	[95% Conf.	Interval]
5.263	97.81	0.0000	3.642	7.607

Displaying mortality rates & survival probabilities





From estimation to modelling: Poisson model

Model the effect of one exposure variable:

- **Rate ratio** = $\frac{\text{Rate in exposed}}{\text{Rate in unexposed}}$

- Rate in exposed = rate in unexposed x **Rate ratio**

- We need unbounded (-∞ to ∞) estimates so we model the log(rate)
- log(rate in exposed) = log(rate in unexposed) + log(**rate ratio**)

- **Assumptions**:
    - Constant rates over time
    - Constant rate ratio over time

From estimation to modelling: Poisson model

Model the effect of one exposure variable:

- ***Hazard* ratio** = $\frac{\textit{Hazard}\ \text{in exposed}}{\textit{Hazard}\ \text{in unexposed}}$

- *Hazard* in exposed = *hazard* in unexposed x **Hazard ratio**

- We need unbounded (-∞ to ∞) estimates so we model the log(*hazard*)
- log(*hazard* in exposed) = log(*hazard* in unexposed) + log(***hazard* ratio**)

- **Assumptions**:
  - Constant *hazards* over time
  - Constant *hazard* ratio over time

Poisson model in R

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

coeftest(prm_died_vimp)
```

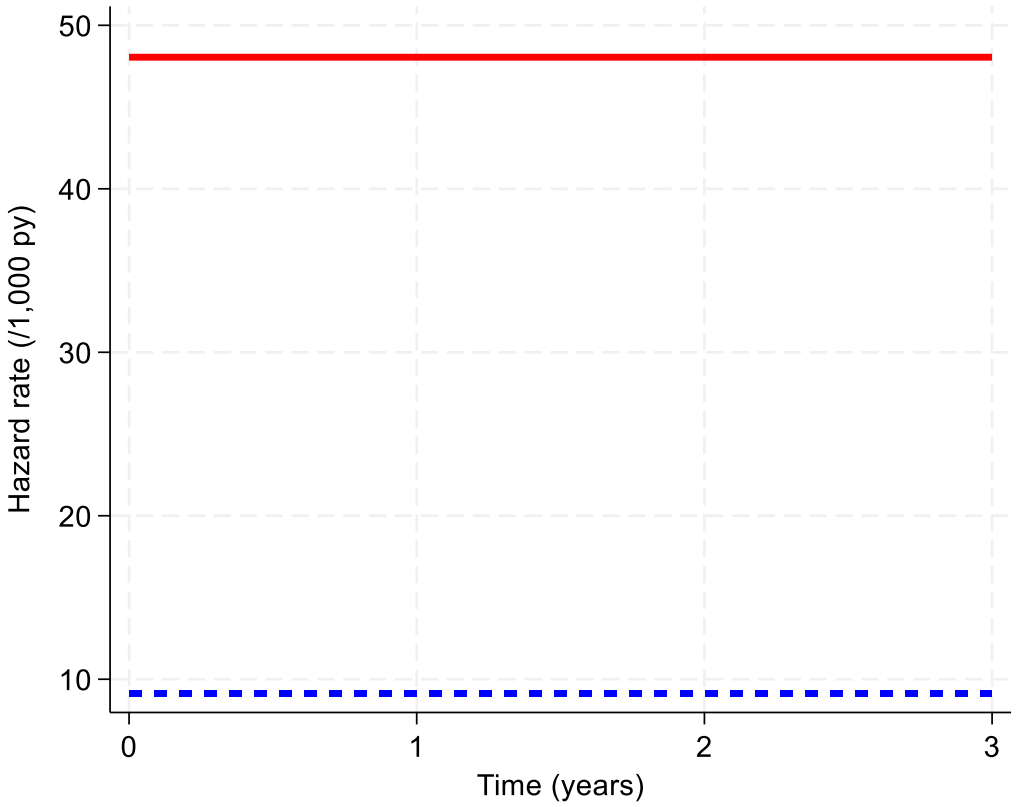
Poisson model in R

	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

HR = 5.26 (95% CI: 3.64 - 7.61; P<0.001)

_cons = 0.00913 (9 / 1000 pyrs) = hazard rate in those visually unimpaired

Graphically



- ***Hazard* ratio =** *Hazard* in exposed / *Hazard* in unexposed
- ***Hazard* ratio =** **5.26**

_cons = 0.00913 (9 / 1000 pyrs)

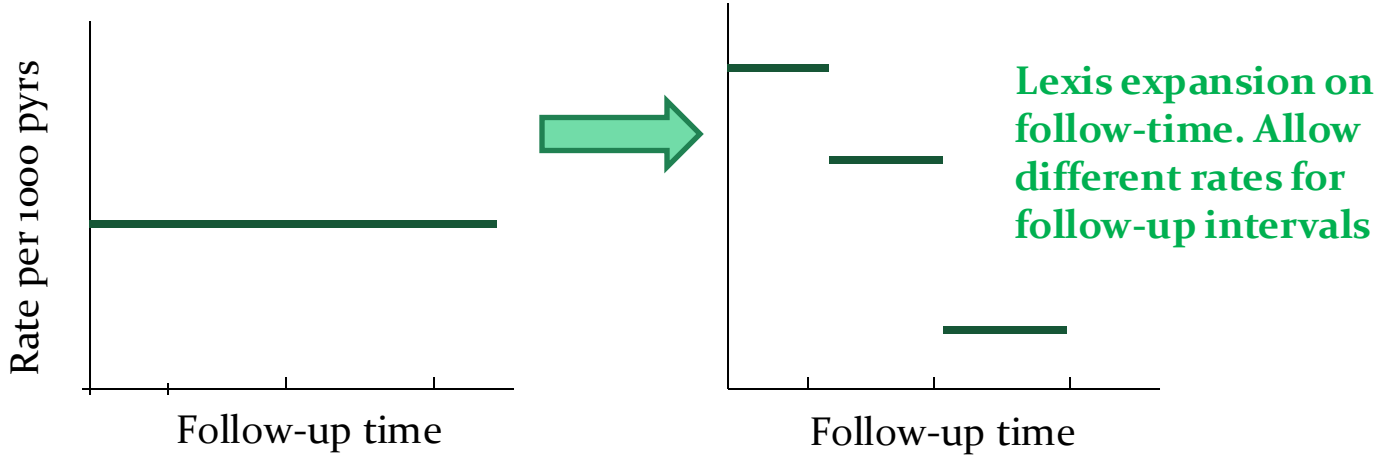
Comments on Poisson regression

Assumed hazard rates constant over follow-up

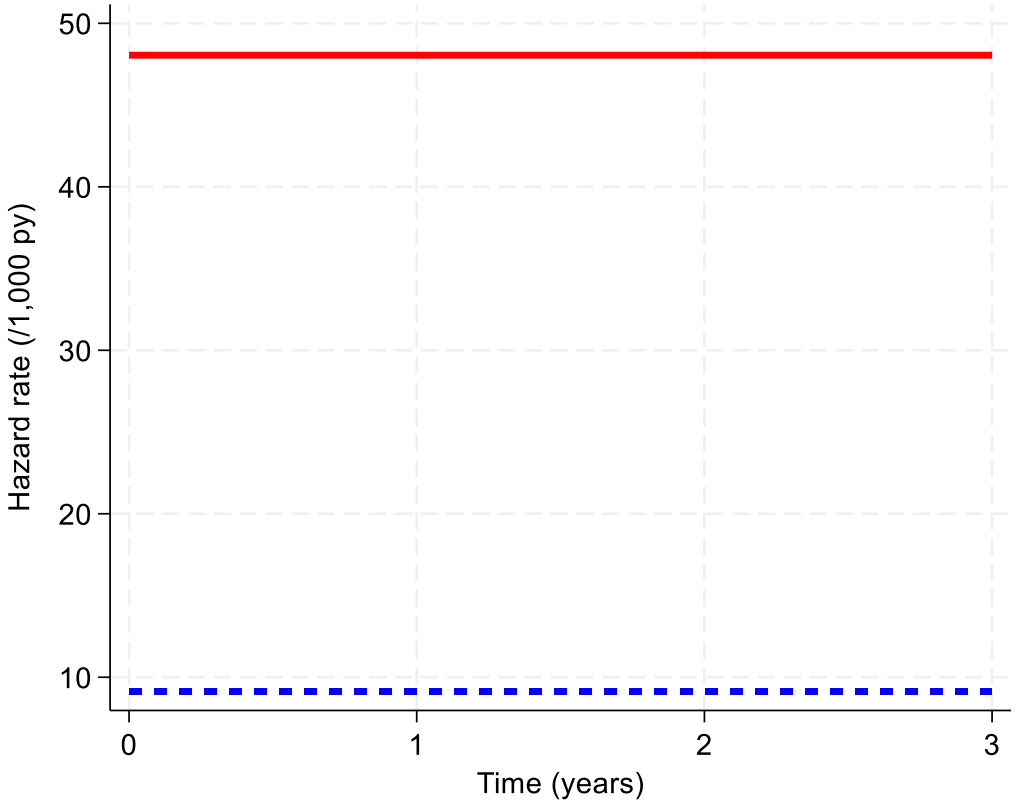
In this study follow-up is short ~ 3yrs, so perhaps a reasonable assumption

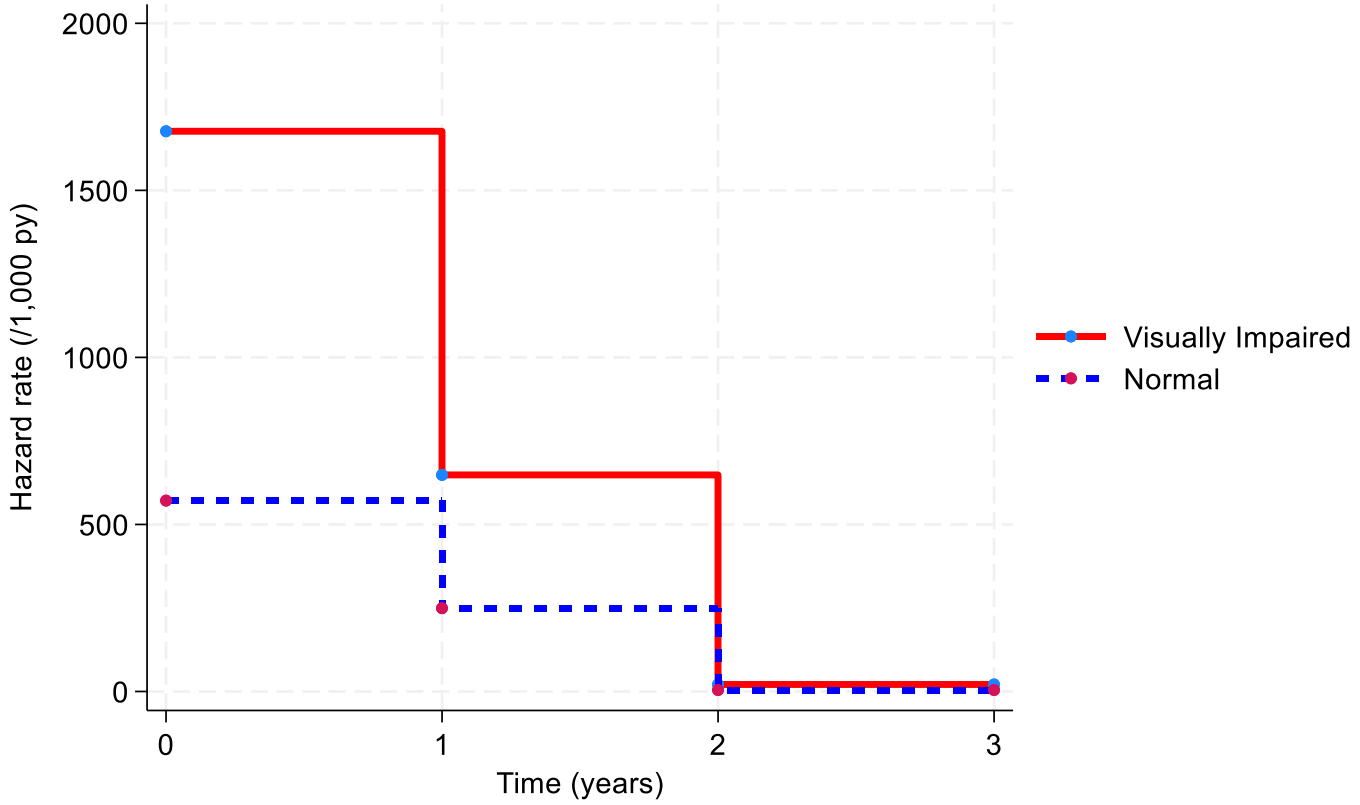
Might not be sensible in other situations (e.g. long cohort study)

If hazard rates not changing too rapidly, can still use Poisson model

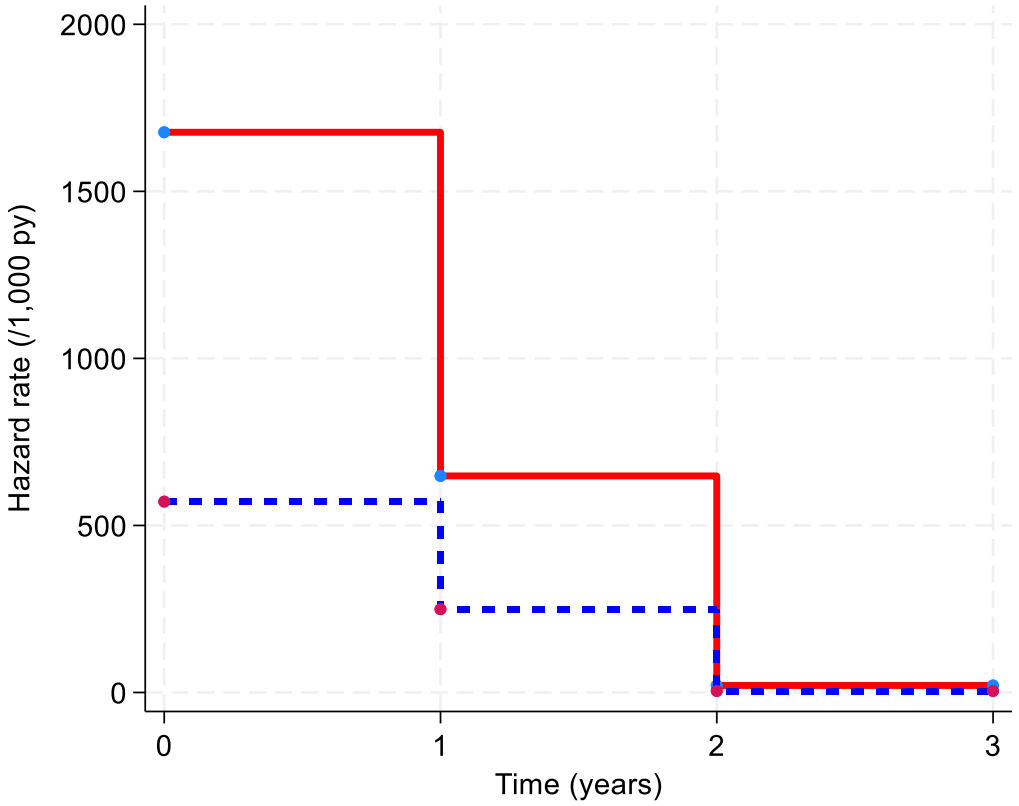


Splitting follow-up time





Splitting follow-up time



- **Assumptions**:
  - *hazards* vary over time
  - Constant *hazard* ratio over time

log(*hazard* exposed) = log(*hazard* unexposed & baseline)+log(*HR* interval k *vs*. baseline)+log(***HR***)

Poisson model with adjusted for fup

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

Poisson model with adjusted for fup

	Haz. ratio	Std. err.	z	P>|z|	[95% conf.	interval]
Visually impaired	**3.260973**	.622376	6.19	0.000	2.243316	4.740281
fup						
1	.411295	.0889747	-4.11	0.000	.2691625	.6284812
2	.0087348	.0019198	-21.57	0.000	.0056776	.0134381
_cons	.5580078	.095078	-3.42	0.001	.3995816	.7792469

Adjusted HR = 3.26 (95% CI: 2.24 - 4.74; P<0.001)

_cons = 0.5580 (558 / 1000 pyrs) = hazard rate for normal vision in first time interval

Same tools as Logistic Regression

- Wald test; Likelihood Ratio Test >> effects of exposures
- Modelling dose response >> ordered categorical variables

Also …

- Adjustment for confounding
- Fitting interactions

Likelihood Ratio Test

- Carry out hypothesis tests using the likelihood ratio test

- Compare the likelihoods from models representing the null & alternative hypotheses: $-2(L_0 - L_1) \sim \chi^2$

where L_0 = log likelihood under Null;

L_1= log likelihood under alternative

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

Likelihood Ratio Test in R

`prm_died_vimp_fup` **>> effect of vimp and follow-up**

	Haz. ratio	Std. err.	z	P>|z|	[95% conf. interval]
Visually impaired	**3.260973**	.622376	6.19	0.000	2.243316 4.740281
fup					
1	.411295	.0889747	-4.11	0.000	.2691625 .6284812
2	.0087348	.0019198	-21.57	0.000	.0056776 .0134381
_cons	.5580078	.095078	-3.42	0.001	.3995816 .7792469

Likelihood Ratio Test in R

lrtest(prm_died_vimp, prm_died_vimp_fup)

```
Likelihood ratio test

Model 1: died ~ vimp + offset(log_p_years)
Model 2: died ~ vimp + factor(fup) + offset(log_p_years)

  #Df  LogLik Df  Chisq Pr(>Chisq)
1   2 -691.49
2   4 -431.24  2 520.51  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Assumption: L1 nested within L2

Assessing for a dose response

- Ordered categorical variable – age group (4 levels)

- Two hypotheses of interest
  - Test for linear trend
  - Test for departures from linearity

Assessing for a dose response

4 categorical ordered age groups: 15-34; 35-54; 55-64; 65+ years

Test for linear trend effect (using the group variable):

```
 prm_died_agegrp_linear >> effect of agegrp as continuous
(1,2,3,4)
```

	Haz. ratio	Std. err.	z	P>|z|	[95% conf.	interval]
agegrp	**2.258283**	.1742578	**10.56**	**0.000**	1.941317	2.627002
_cons	.0072917	.0010179	-35.25	0.000	.0055464	.0095863

Assessing for a dose response

Test for departure from linearity:

```
prm_died_agegrp >> effect of agegrp category (15-34; 35-54; 55-64; 65+ years)
```

_t	Haz. ratio	Std. err.	z	P>|z|	[95% conf.	interval]
agegrp						
35-54	2.443815	.5630934	3.88	0.000	1.555744	3.838827
55-64	6.109201	1.624629	6.81	0.000	3.627631	10.28835
65+	11.31935	2.852569	9.63	0.000	6.907359	18.54946
_cons	.0068988	.0012391	-27.71	0.000	.0048517	.0098097

Assessing for a dose response

Test for departure from linearity:

```
lrtest(prm_died_agegrp_linear, prm_died_agegrp)

Likelihood ratio test

Model 1: died ~ cont_age + offset(log_p_years)
Model 2: died ~ agegrp + offset(log_p_years)

  #Df  LogLik Df  Chisq     Pr(>Chisq)
1   2 -666.99
2   4 -666.78  2 0.4161       0.8122 >> no evidence for departure from linearity
```

How to present results?

Good practice to show:

- The disease burden, overall and by exposure variable:
  - no. of events,
  - person-time
  - rates (per 100 pyrs, for example)
  - RRs, 95% CIs

- For variables with >2 levels:
  - show the overall (single) pvalue (from the likelihood ratio test) rather than pvalues for each RR

Example

	Deaths/pyrs	Rate per 1,000 pyrs	RR	95% CI	P-value
Visual impairement					
Normal vision	97/10,625	9.1	1		<0.001
Visually impaired	40/832	48.1	5.3	(3.6-7.6)	
Sex					
Male	70/5,401	13.0	1		0.4
Female	67/6,057	11.1	0.9	(0.6-1.2)	
Age					
15-34	31/6300	4.9	1		<0.001
35-54	48/3900	12.4	2.4	(1.5-3.8)	
55-64	26/810	32.1	6.1	(3.6-10.3)	
65+	32/491	65.2	11.3	(6.9-18.5)	

Summary

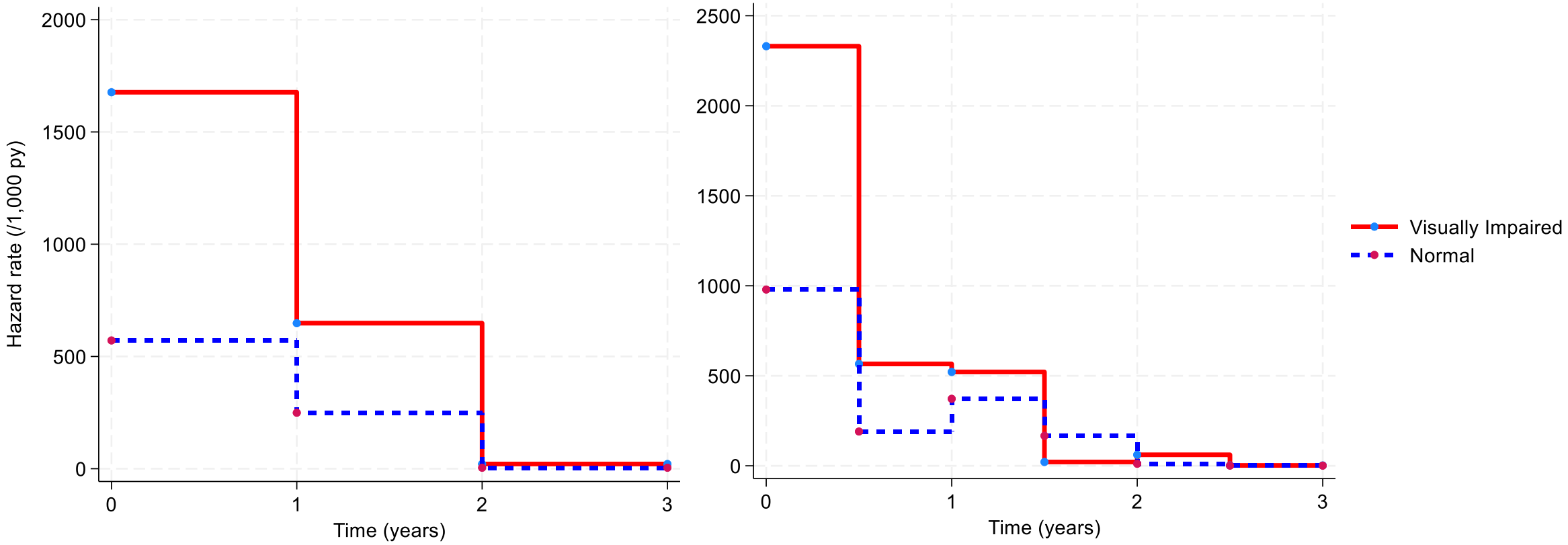
- Poisson model is used to obtain rates and rate ratios
- Model can be used to examine
  - Significance of variables (Wald, LRT)
  - Dose response
  - Confounding
  - Effect modification

- Assumptions of the Poisson model:
  - Constant hazard, but can be relaxed
  - Constant HR

Poisson *vs.* Cox

Sometimes the rate may change very rapidly with time,

Poisson models become difficult to fit



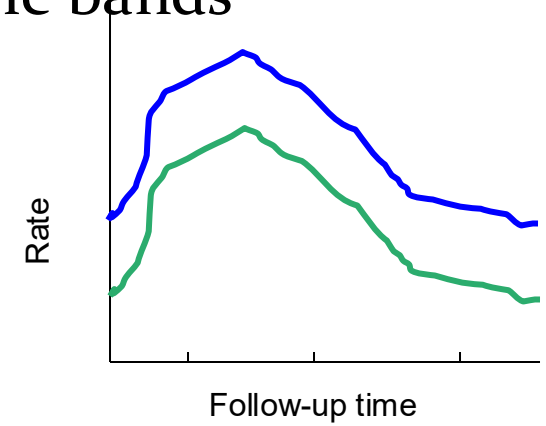
Cox model

- Analysing the time elapsed between entry to the study and the event of interest

- Particularly useful
  - where rates *vary rapidly with time*
  - where high proportion of individuals experience the event

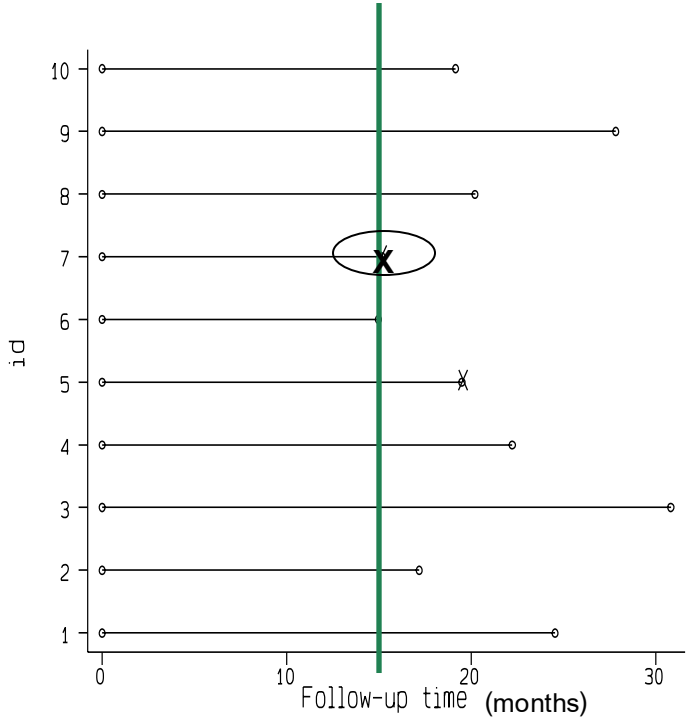
- Estimate the baseline hazard for smallest intervals possible:
  each interval only contains one event >> non-parametric estimation
  Allows *baseline hazard rates* to vary with (infinitesimal) time bands

- Assumption of constant hazard ratios: *proportional hazards*



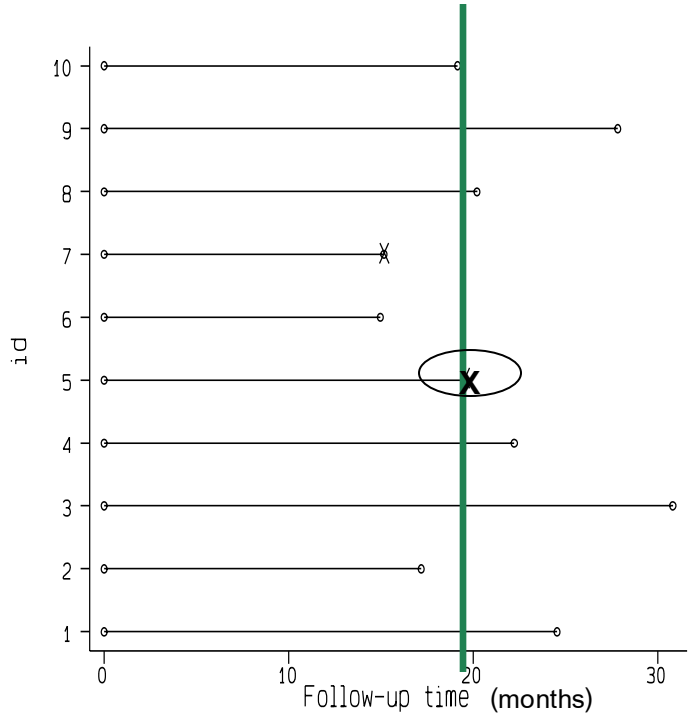
Cox model – risk sets

- Creates *risk sets* at each failure (outcome event)
- For example: at 15 months id=7 dies
- The risk set is everyone who is alive <u>at that follow-up time</u>
- Compares exposure status of event with everyone in risk set



Cox model – risk sets (2)

- Next death occurred at ~19.5 years, id=5
- The risk set is everyone who is alive at that follow-up time
- Compares exposure status of event with everyone in risk set



Risk sets are "matched on follow-up time"

Cox model in R

```
cox_vimp <- coxph(Surv(p_years,died) ~ vimp, data = mortality)
summary(cox_vimp)
```

```
------------------------------------------------------------------------------
                    | Haz. ratio   Std. err.      z    P>|z|     [95% conf. interval]
--------------------+---------------------------------------------------------
Visually impaired   |   4.529811   .8512425     8.04   0.000     3.134168    6.546934
------------------------------------------------------------------------------
```

- HR=4.53 (95% CI: 3.13, 6.55)

N.B Because risk sets were based ("matched") on time from entry to study, the HR has been controlled for any effect of follow-up time (time from entry to study)

General comments

Poisson model can be used to obtain rates (hazards) and rate (hazard) ratios

- Assumes constant rates
- Can carry out hypothesis tests (LRT, Wald), control for confounders, assess effect modification and test for linear trend
- Model is useful provided rates do not vary rapidly with time

Cox model examines time to event & provides rate (hazard) ratio

- Useful if rates (hazard) vary rapidly with time
- Can carry out hypothesis tests, control for confounding, assess effect modification, and test for linear trend
- Assumes proportional hazards (constant rate ratio)
  - Needs to be assessed (graphically and using a hypothesis test)

Thank you!

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE