# Biostatistics III: Survival analysis for epidemiologists

Department of Medical Epidemiology and Biostatistics
Karolinska Institutet
Stockholm, Sweden

`http://www.biostat3.net/`

# Extra lecture on modelling

- This lecture covers general theory of modelling and effect parameterisation (not specific to survival analysis).

- It will cover
  - Categorical and continuous effects
  - Main effects model: Interpretation and tests of main effects
  - Interaction models: Interpretation and tests of interaction effects
  - Parameterisation of interaction effects to obtain the group comparison of interest

- As illustrative example, we will use the Poisson regression model for the all-cause mortality rate, assuming constant (average) rates over time. So we will ignore time for now.

# Categorical exposure variables

- We will use the melanoma data set investigating age differences in all-cause mortality. The data are stset on surv_mm and status=1,2.

- Age at diagnosis is measured in years in the variable age.

- The variable age3 (created below) has 3 levels. Here are the crude rates and HRs.

```
. egen age3=cut(age),at(0,60,75,100)
```

| Age | age3 | Rate/1000py | HR |
|---|---|---|---|
| 0-59 | 0 | 32.8 | 1.0 |
| 60-74 | 60 | 81.8 | 81.8/32.8=2.49 |
| 75+ | 75 | 191.4 | 191.4/32.8=5.84 |

- To include age3 in regression commands we need to use indicator variables for the 3 levels.

| age3 | $X1$ | $X2$ | $X3$ |
|---|---|---|---|
| 0 | 1 | 0 | 0 |
| 60 | 0 | 1 | 0 |
| 75 | 0 | 0 | 1 |

- We create dummy variables X1, X2, X3 for exposure age3.

```
. tabulate age3, generate(X)
```

- We fit a Poisson regression model, and set exposure level 0-59 as reference by omitting X1 from the model.

```
. stset surv_mm, failure(status==1,2) scale(12)

. poisson _d X2 X3, exposure(surv_mm) irr

          _d |        IRR   [95% Conf. Interval]
-------------+--------------------------------
          X2 |   2.494544  2.294322   2.712240
          X3 |   5.835393  5.335520   6.382098
       _cons |   .0027329  .0025756   .0028997
 ln(surv_mm) |          1
```

- The variable $(X1)$ that indicates the category with the lowest age (0-59) is omitted, meaning this is the reference category.

- The irr option will exponentiate the parameters to rate ratios.

- Note that Stata will display IRR even though we have estimated a hazard ratio for all-cause mortality.

- In terms of the parameters

$$\ln(\lambda) = \beta_0 + \beta_2 X_2 + \beta_3 X_3$$

| age | X | ln(rate) | rate |
|---|---|---|---|
| 0-59 | $X_1$ | $\beta_0$ | $\exp(\beta_0)$ |
| 60-74 | $X_2$ | $\beta_0 + \beta_2$ | $\exp(\beta_0 + \beta_2) = \exp(\beta_0)\exp(\beta_2)$ |
| 75+ | $X_3$ | $\beta_0 + \beta_3$ | $\exp(\beta_0 + \beta_3) = \exp(\beta_0)\exp(\beta_3)$ |

$$\text{IRR}_{X_2/X_1} = \exp(\beta_2) = \frac{\exp(\beta_0)\exp(\beta_2)}{\exp(\beta_0)}$$

$$\text{IRR}_{X_3/X_1} = \exp(\beta_3) = \frac{\exp(\beta_0)\exp(\beta_3)}{\exp(\beta_0)}$$

# Automatic generation of indicators using factor variables

```
. poisson _d i.age3, exposure(surv_mm) irr
-------------------------------------------
      _d |       IRR    [95% Conf. Interval]
---------+---------------------------------
    age3 |
     60  |   2.494544  2.294322    2.712240
     75  |   5.835393  5.33552     6.382098
   _cons |   .0027329  .0025756    .0028997
```

- i. tells Stata that age3 should be treated as a categorical variable
- The baseline is, by default, the first level (0-59), but this can be changed to (say) the third level (75+) with

```
. poisson _d ib75.age3, exposure(surv_mm) irr
```

# Metric (continuous) exposure variables

- The effect of age on the mortality rate, when age is measured as a continuous variable can be estimated like this:

```
. poisson _d age , exposure(surv_mm) irr
----------------------------------------
      _d |       IRR     [95% Conf. Interval]
-------+--------------------------------
     age |    1.048671   1.04609    1.051258
```

- For each 1 unit increase in age, the mortality rate is increased by 4.9%.

- This model makes the strong assumption of linearity of the effect, which may not be plausible.

- A 1 unit increase may not be clinically relevant. We may want to rescale the continuous variable.

- To get the HR for an increase of, say, 10-year intervals:

```
. gen age10=age/10
. poisson _d age10, exposure(surv_mm) irr
----------------------------------------
       _d |        IRR  [95% Conf. Interval]
----------+-----------------------------
    age10 |   1.608398  1.56925    1.648522
```

- The estimated IRR is $1.048671^{10} = 1.6084$. That is, for each 10 unit increase in age, we estimate that the mortality rate is increased by 61%.

- Again, this model makes the assumption on linearity of the effect.

## The main effects model — constant effect over strata

- If the true effect of exposure does not vary across strata of another variable we can use a main effects model.
- For example, if the effect of age is the same in both sexes, we can estimate an age effect that is the same for both men and women.
- Thus, if the estimates of age differ only randomly over sex level, we can consider a model in which the true effect of age is constant over sex, i.e. no interaction.
- This allows us to combine the information from different strata to yield a single estimate of exposure effect.
- This combined estimate of the effect we call *the main effect*, which is then *controlled for* the stratifying (confounding) variable(s).

- The effect of exposure is assumed to be the same in all levels of other variables in a main effects model.

- Statistical tests for the presence of confounding are not available, although there are statistical tests for effect modification.

# Main effects model using Poisson regression

```
. poisson _d ib1.sex ib0.age3, exposure(surv_mm) irr baselevels
------------------------------------------------------------------
          _d |        IRR        z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------
         sex |
       Male  |          1  (base)
     Female  |   .6140348   -13.36   0.000     .5716519    .6595599
             |
        age3 |
          0  |          1  (base)
         60  |   2.499009    21.45   0.000     2.298428    2.717096
         75  |   6.184747    39.69   0.000     5.652568    6.767029
             |
       _cons |   .0034875  -165.36   0.000     .0032612    .0037294
```

- The `poisson` regression command makes no distinction between the exposure variable and the control variable.

- The first hazard ratio reported is the effect of `sex` controlled for `age3`, and the next two are the effects of `age3` controlled for `sex`.
- The p-value of the parameter for sex (HR=0.614) is $<0.001$. This means that there is a statistically significant association between sex and all-cause mortality, when adjusting for age.
- To test the effect of age, we need to test both age parameters jointly.

```
. testparm 60.age3 75.age3

 ( 1)  [_d]60.age3 = 0
 ( 2)  [_d]75.age3 = 0

           chi2(  2) = 1586.33
         Prob > chi2 =    0.0000
```

- The effect of age is statistically significant (p-value$<0.001$) when adjusting for sex.

# Models and parameters in Poisson regression

- In the Poisson regression model we estimated 4 parameters. One parameter (the intercept) is a log rate and the other three are log hazard ratios.

- The model is $\ln(\lambda) = \beta_0 + \beta_1\text{female} + \beta_2\text{age60} + \beta_3\text{age75}$

- Note that female, age60 and age75 are dummy variables coded 0/1 (similar to X1, X2, X3).

- The parameters are

$$
\begin{aligned}
\exp(\beta_0) &= \text{rate at reference level of all covariates, i.e. men, 0-59 years} \\
\exp(\beta_1) &= \text{rate ratio (comparing women vs men)} \\
\exp(\beta_2) &= \text{rate ratio (comparing 60-74 vs 0-59)} \\
\exp(\beta_3) &= \text{rate ratio (comparing 75+ vs 0-59)}
\end{aligned}
$$

# Parameters estimates with and without the `irr` option

```
. poisson _d ib0.female ib0.age3, exposure(surv_mm)
          _d |      Coef.   [95% Conf. Interval]
-------------+--------------------------------------
     Female  |  -.4877037    -.5592251   -.4161824
             |
         60  |   .9158944    .8322253    .9995635
         75  |   1.822086    1.73211     1.912062
       _cons |  -5.658582    -5.725652   -5.591511

. poisson _d ib0.female ib0.age3, exposure(surv_mm) irr
          _d |        IRR   [95% Conf. Interval]
-------------+-------------------------------------
     Female  |   .6140348   .5716519    .6595599
             |
         60  |   2.499009   2.298428    2.717096
         75  |   6.184747   5.652568    6.767029
       _cons |   .0034875   .0032612    .0037294
```

- The irr option exponentiates the coefficients.

- From the model, the estimated rate for each combination of explanatory variables can be formulated as a function of the baseline rate $\lambda$ and the three incidence rate ratios. The baseline is the reference group of all variables (men, 0-59 years).

- The model is $\ln(\lambda) = \beta_0 + \beta_1\text{female} + \beta_2\text{age60} + \beta_3\text{age75}$

- Which on the rate scale is $\lambda = \exp(\beta_0 + \beta_1\text{female} + \beta_2\text{age60} + \beta_3\text{age75})$

- These are the rates, $\lambda$:

| | female=0 (men) | female=1 (women) |
|---|---|---|
| age=0-59 | $\exp(\beta_0)$ | $\exp(\beta_0 + \beta_1)$ |
| age=60-74 | $\exp(\beta_0 + \beta_2)$ | $\exp(\beta_0 + \beta_1 + \beta_2)$ |
| age=75+ | $\exp(\beta_0 + \beta_3)$ | $\exp(\beta_0 + \beta_1 + \beta_3)$ |

- The estimated incidence rate for men 0-59y (ref group) is $\exp(-5.658582) = 0.00349$ events/person-year.

- The estimated incidence rate for women 60-74y is $\exp(\beta_0 + \beta_1 + \beta_2) = \exp(\beta_0) \times \exp(\beta_1) \times \exp(\beta_2) = 0.0034875 \times 0.6140348 \times 2.499009 = 0.00535$ events/person-year.

- To get the HRs (with reference group 'men 0-59y') we simply divide all cells with the baseline rate $\lambda = \exp(\beta_0)$

| | female=0 (men) | female=1 (women) |
|---|---|---|
| age=0-59 | $\exp(\beta_0)/\exp(\beta_0) = 1$ | $\exp(\beta_0 + \beta_1)/\exp(\beta_0) = \exp(\beta_1)$ |
| age=60-74 | $\exp(\beta_0 + \beta_2)/\exp(\beta_0) = \exp(\beta_2)$ | $\exp(\beta_0 + \beta_1 + \beta_2)/\exp(\beta_0) = \exp(\beta_1 + \beta_2)$ |
| age=75+ | $\exp(\beta_0 + \beta_3)/\exp(\beta_0) = \exp(\beta_3)$ | $\exp(\beta_0 + \beta_1 + \beta_3)/\exp(\beta_0) = \exp(\beta_1 + \beta_3)$ |

- If we put the estimates from the Stata output into our table of parameters, we get the HRs

| | female=0 (men) | female=1 (women) |
|---|---|---|
| age=0-59 | $1.0$ | $0.61$ |
| age=60-74 | $2.50$ | $0.61 \times 2.50 = 1.53$ |
| age=75+ | $6.18$ | $0.61 \times 6.16 = 3.80$ |

- Compared to men aged 0-59y, the women aged 60-74y have a 53% higher mortality (HR=1.53), which is a combined effect of age and sex.

- Effect of being woman vs man is 0.61 (39% lower mortality).

- Effect of being 60-74 vs 0-59 is 2.50 (250% higher, or 2.5 times higher).

- Effect of being 75+ vs 0-59 is 6.18 (618% higher, or 6.2 times higher).

- The effects of age are adjusted for sex, meaning that we are comparing age effects within levels of sex. The effect of age (HRs 2.50 and 6.18) is the same for both men and women, and is the averaged age effect across levels of sex.

- Similarly, the effect of sex is adjusted for age, meaning that we are comparing the sex effect within levels of age. The effect of sex (HR 0.61) is the same in all age groups, and is averaged across age groups.

- If we want to get the effect of sex separately in all levels of age, we simply change the reference rate to be men in all age levels.

| | female=0 (men) | female=1 (women) |
|---|---|---|
| age=0-59 | $\exp(\beta_0)/\exp(\beta_0) = 1$ | $\exp(\beta_0+\beta_1)/\exp(\beta_0) = \exp(\beta_1)$ |
| age=60-74 | $\exp(\beta_0+\beta_2)/\exp(\beta_0+\beta_2) = 1$ | $\exp(\beta_0+\beta_1+\beta_2)/\exp(\beta_0+\beta_2) = \exp(\beta_1)$ |
| age=75+ | $\exp(\beta_0+\beta_3)/\exp(\beta_0+\beta_3) = 1$ | $\exp(\beta_0+\beta_1+\beta_3)/\exp(\beta_0+\beta_3) = \exp(\beta_1)$ |

- We see that the main effect models give the same effect of sex, $\exp(\beta_1)$, regardless of age level.

| | female=0 (men) | female=1 (women) |
|---|---|---|
| age=0-59 | $1.0$ | $\exp(\beta_1)$ |
| age=60-74 | $1.0$ | $\exp(\beta_1)$ |
| age=75 | $1.0$ | $\exp(\beta_1)$ |

- This is what a main effect model assumes, i.e. the effect of exposure is the same in all levels of another variable.

- Similarly, if we want to get the effect of age in both levels of sex, we change the reference rate to 0-59 in both men and women.

| | female=0 (men) | female=1 (women) |
|---|---|---|
| age=0-59 | $\exp(\beta_0)/\exp(\beta_0) = 1$ | $\exp(\beta_0 + \beta_1)/\exp(\beta_0 + \beta_1) = 1$ |
| age=60-74 | $\exp(\beta_0 + \beta_2)/\exp(\beta_0) = \exp(\beta_2)$ | $\exp(\beta_0 + \beta_1 + \beta_2)/\exp(\beta_0 + \beta_1) = \exp(\beta_2)$ |
| age=75+ | $\exp(\beta_0 + \beta_3)/\exp(\beta_0) = \exp(\beta_3)$ | $\exp(\beta_0 + \beta_1 + \beta_3)/\exp(\beta_0 + \beta_1) = \exp(\beta_3)$ |

- Again, the main effects of age are the same, $\exp(\beta_2)$ and $\exp(\beta_3)$, regardless of sex level.

| | female=0 (men) | female=1 (women) |
|---|---|---|
| age=0-59 | $1.0$ | $1.0$ |
| age=60-74 | $\exp(\beta_2)$ | $\exp(\beta_2)$ |
| age=75+ | $\exp(\beta_3)$ | $\exp(\beta_3)$ |

# Effect modification

- If the true effect of exposure is not the same, but varies across strata of another variable there is said to be ‘effect modification’ (interaction).

- The exposure effect cannot then be represented by one HR for all levels of the other variable.

- For example, the effect of age may depend on sex, so that men have a greater effect of age than women.

- Then we say that the effect of age is *modified by sex*. There is an interaction between age and sex.

- Does `sex` modify the effect of `age`? If we estimate, from mortality rates, the HR of women vs men separately in all age groups we get

| Age | Rate men | Rate women | HR (women vs men) |
|---|---|---|---|
| 0-59 | 46.7 | 21.9 | 21.9/46.7=0.469 |
| 60-74 | 98.0 | 69.3 | 69.3/98.0=0.707 |
| 75+ | 235.9 | 170.0 | 170.0/235.9=0.721 |

- The hazard ratios compare mortality in women to men within each age category.

- If the effect of age is not modified by sex then we would expect these to be similar.

- In the previous main effect model, we assumed (and estimated) the effect to be 0.61 in all age groups. Is that plausible?

## Interaction model using Poisson regression

- If we want to estimate separate effects of sex in age levels, then four parameters are not enough.

- We have six combinations of age and sex, hence we need six parameters to estimate six different rates.

- This is easily done by including interaction terms in the Poisson model.

- The interaction model is $\ln(\lambda) =$ $\beta_0 + \beta_1$female $+ \beta_2$age60 $+ \beta_3$age75 $+ \beta_4$female x age60 $+ \beta_5$female x age75

- The parameters are

$$
\begin{aligned}
\exp(\beta_0) &= \text{rate at reference level of all covariates, i.e. men 0-59} \\
\exp(\beta_1) &= \text{rate ratio (comparing women vs men among 0-59)} \\
\exp(\beta_2) &= \text{rate ratio (comparing 60-74 vs 0-59 among men)} \\
\exp(\beta_3) &= \text{rate ratio (comparing 75+ vs 0-59 among men)} \\
\exp(\beta_4) &= \text{interaction term (excess rate for 60-74 vs 0-59 among women)} \\
\exp(\beta_5) &= \text{interaction term (excess rate for 75+ vs 0-59 among women)}
\end{aligned}
$$

```
. poisson _d ib0.female##ib0.age3, exposure(surv_mm) irr
           _d |        IRR   [95% Conf  Int]
--------------+------------------------
      Female  |    .468009   .4140604    .5289866
              |
         age3 |
          60  |   2.097968   1.877825    2.343918
          75  |   5.047386   4.432543    5.747514
              |
  female#age3 |
   Female#60  |   1.509675   1.273345    1.789869
   Female#75  |   1.539727   1.281208     1.85041
              |
        _cons |   .0038943   .0036133    .0041971
```

- 0.468 is the effect of `female` when age3 is at its first level.
- 2.097 and 5.047 are the effects of age3 when `female` is at its first level.
- 1.509 and 1.539 are the interactions between `female` and age3.

- 0.00389 is the baseline rate in the reference level of both female and age3.

# Parameters for the interaction model

- The *fitted rates* ($\lambda$) from the interaction model can be expressed in terms of the model parameters as follows

| | female=0 (men) | female=1 (women) |
|---|---|---|
| age=0-59 | $\exp(\beta_0)$ | $\exp(\beta_0 + \beta_1)$ |
| age=60-74 | $\exp(\beta_0 + \beta_2)$ | $\exp(\beta_0 + \beta_1 + \beta_2 + \beta_4)$ |
| age=75+ | $\exp(\beta_0 + \beta_3)$ | $\exp(\beta_0 + \beta_1 + \beta_3 + \beta_5)$ |

- $\exp(\beta_4)$ and $\exp(\beta_5)$ are the interaction parameters. They measure deviations from the hypothesis of main effect of sex in all age categories.

- The model is $\ln(\lambda) =$
  $\beta_0 + \beta_1$female $+ \beta_2$age60 $+ \beta_3$age75 $+ \beta_4$female x age60 $+ \beta_5$female x age75

- If we wish to tabulate the HRs, using the men aged 0-59 as reference group, then we simply divided all cells with the reference rate.

| | female=0 (men) | female=1 (women) |
|---|---|---|
| 0-59 | $\exp(\beta_0)/\exp(\beta_0) = 1$ | $\exp(\beta_0 + \beta_1)/\exp(\beta_0) = \exp(\beta_1)$ |
| 60-74 | $\exp(\beta_0 + \beta_2)/\exp(\beta_0) = \exp(\beta_2)$ | $\exp(\beta_0 + \beta_1 + \beta_2 + \beta_4)/\exp(\beta_0) = \exp(\beta_1 + \beta_2 + \beta_4)$ |
| 75+ | $\exp(\beta_0 + \beta_3)/\exp(\beta_0) = \exp(\beta_3)$ | $\exp(\beta_0 + \beta_1 + \beta_3 + \beta_5)/\exp(\beta_0) = \exp(\beta_1 + \beta_3 + \beta_5)$ |

- If we add the Stata output into the table of HRs we get

| | female=0 (men) | female=1 (women) |
|---|---|---|
| age=0-59 | $1.0$ | $0.468$ |
| age=60-74 | $2.098$ | $0.468 \times 2.098 \times 1.510 = 1.48$ |
| age=75+ | $5.047$ | $0.468 \times 5.047 \times 1.540 = 3.64$ |

- Compare to the estimates from main effects model

| | female=0 (men) | female=1 (women) |
|---|---|---|
| age=0-59 | $1.0$ | $0.61$ |
| age=60-74 | $2.50$ | $0.61 \times 2.50 = 1.53$ |
| age=75+ | $6.18$ | $0.61 \times 6.16 = 3.80$ |

# Testing for interaction

- A test of interaction is simply to test if the excess (1.510 and 1.540) is equal to 1 (or coefficients equal 0 on the log scale).

```
. poisson _d ib0.female##ib0.age3, exposure(surv_mm) irr
------------------------------------------------------------------------------
          _d |        IRR   Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     Female  |    .468009   .0292453   -12.15   0.000     .4140604    .5289866
             |
        age3 |
         60  |   2.097968   .1186601    13.10   0.000     1.877825    2.343918
         75  |   5.047386   .3345166    24.43   0.000     4.432543    5.747514
             |
 female#age3 |
  Female#60  |   1.509675   .1311343     4.74   0.000     1.273345    1.789869
  Female#75  |   1.539727   .1443928     4.60   0.000     1.281208     1.85041
             |
       _cons |   .0038943   .0001488  -145.21   0.000     .0036133    .0041971
------------------------------------------------------------------------------
```

```
. testparm 1.female#60.age3 1.female#75.age3

 ( 1)  [_d]1.female#60.age3 = 0
 ( 2)  [_d]1.female#75.age3 = 0

           chi2(  2) =   29.53
         Prob > chi2 =    0.0000
```

- There is evidence of a statistically significant interaction (p-value $<0.0001$).

- This is a so-called Wald test, which approximates the likelihood ratio test. We could also use a likelihood ratio test, where we compare the log-likelihoods from the main effects model and the interaction model.

## Reparameterising the model to directly estimate the effect of exposure in each level of another variable

- We are often interested in the effect of the exposure (women vs men) for each level of the modifier (age group).

- We can divide the rates with different reference rates to obtain the effect of sex in each level of age.

| | female=0 (men) | female=1 (women) |
|---|---|---|
| 0-59 | $\exp(\beta_0)/\exp(\beta_0) = 1$ | $\exp(\beta_0 + \beta_1)/\exp(\beta_0) = \exp(\beta_1)$ |
| 60-74 | $\exp(\beta_0 + \beta_2)/\exp(\beta_0 + \beta_2) = 1$ | $\exp(\beta_0 + \beta_1 + \beta_2 + \beta_4)/\exp(\beta_0 + \beta_2) = \exp(\beta_1 + \beta_4)$ |
| 75+ | $\exp(\beta_0 + \beta_3)/\exp(\beta_0 + \beta_3) = 1$ | $\exp(\beta_0 + \beta_1 + \beta_3 + \beta_5)/\exp(\beta_0 + \beta_3) = \exp(\beta_1 + \beta_5)$ |

- This yields the following HRs:

| | female=0 (men) | female=1 (women) |
|---|---|---|
| age=0-59 | $1.0$ | $\exp(\beta_1)$ |
| age=60-74 | $1.0$ | $\exp(\beta_1 + \beta_4)$ |
| age=75+ | $1.0$ | $\exp(\beta_1 + \beta_5)$ |

- We can reparameterise the model in Stata to directly estimate parameters of sex, one for each age group. We use the # notation instead of ##.

# How to make Stata produce stratified effects

- Instead of just one baseline rate, you need three baseline rates, $\lambda$ (one for each age group). In addition, for each age group, you need an HR for the sex effect (three hazard ratios).

```
. poisson _d ibn.age3 ib0.female#i.age3, exposure(surv_mm) irr nocons
-------------------------------------------------
          _d |        IRR    [95% conf. interval]
-------------+-----------------------------------
        age3 |
          0  |   .0038943    .0036133    .0041971
         60  |   .0081701    .0075288    .0088659
         75  |   .0196559    .0176765    .0218569
             |
 female#age3 |
   Female# 0 |    .468009   .4140604    .5289866
   Female#60 |   .7065416   .6277414    .7952336
   Female#75 |   .7206063   .6283156    .8264532
```

- Note that this is the same interaction model; there are still 6 parameters and the fitted rates are identical. It is just that the 6 parameters in this model have a different interpretation.

- The log-likelihood for this model is the same as the previous interaction model. This is because we are fitting the exact same model, but with different parameterisation.

- Technical note: The ibn. factor-variable operator specifies that a categorical variable should be treated as if it has no base, or, in other words, that all levels of the categorical variable are to be included in the parameters. This means that there is no constant (also: nocons option must be added).

## Effects of exposure within each stratum of the modifier

- If we insert the Stata output in our previous table, we get

| | female=0 (men) | female=1 (women) |
|---|---|---|
| age=0-59 | 1.0 | 0.468 |
| age=60-74 | 1.0 | 0.707 |
| age=75+ | 1.0 | 0.721 |

- The stratum-specific HRs are different, there is evidence of interaction. In particular, the HR for 0-59 is much lower than those for 60-74 and 75+.

- Effect of sex is 0.468 for 0-59, 0.468x1.510=0.707 for 60-74, and 0.468x1.540=0.721 for 75+. (There may be rounding errors.)

- Compare this to the main effects model: 0.61 for all age groups.

- Similarly, we can reparameterise the models to show effect of age in levels of sex.

- This yields the following HRs

| | female=0 (men) | female=1 (women) |
|---|---|---|
| age=0-59 | $1.0$ | $1.0$ |
| age=60-74 | $\exp(\beta_2)$ | $\exp(\beta_2 + \beta_4)$ |
| age=75+ | $\exp(\beta_3)$ | $\exp(\beta_3 + \beta_5)$ |

```
. poisson _d ibn.female ib0.age3#i.female, exp(surv_mm) irr nocons
------------------------------------------------
          _d |        IRR   [95% conf. interval]
-------------+----------------------------------
      female |
       Male  |   .0038943   .0036133    .0041971
     Female  |   .0018226   .0016542     .002008

 age3#female |
    60#Male  |   2.097968   1.877825    2.343918
  60#Female  |   3.167250   2.783343    3.604109
    75#Male  |   5.047386   4.432543    5.747514
  75#Female  |   7.771598   6.823946    8.850852
------------------------------------------------
```

- If we insert the Stata output in our previous table, we get

| | female=0 (men) | female=1 (women) |
|---|---|---|
| age=0-59 | 1.0 | 1.0 |
| age=60-74 | 2.098 | 3.167 |
| age=75+ | 5.047 | 7.772 |

- Effect of being 60-74 vs 0-59 is 2.10 for men, whereas it is 3.17 for women.
- Compare this to the main effects model: 2.50 for both sexes.
- Effect of being 75+ vs 0-59 is 5.05 for men, whereas it is 7.77 for women.
- Compare this to the main effects model: 6.18 for both sexes.

# Linear combinations of parameters

- As an alternative to reparameterising the interaction model with # we can use the Stata `lincom()` command to estimate the effect of exposure within each level of the modifier together with confidence intervals. Here again is the interaction model with the default parameterisation.

```
.poisson _d ib0.female##ib0.age3, exposure(surv_mm) irr
------------------------------------------------------------------------------
          _d |        IRR   Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      female |
     Female  |    .468009   .0292453   -12.15   0.000     .4140604    .5289866
             |
        age3 |
         60  |   2.097968   .1186601    13.10   0.000     1.877825    2.343918
         75  |   5.047386   .3345166    24.43   0.000     4.432543    5.747514
             |
 female#age3 |
  Female#60  |   1.509675   .1311343     4.74   0.000     1.273345    1.789869
  Female#75  |   1.539727   .1443928     4.60   0.000     1.281208     1.85041
             |
       _cons |   .0038943   .0001488  -145.21   0.000     .0036133    .0041971
```

- The effect of sex for age 0-59 is $0.468$. We now estimate the effect of sex for the other two categories of age.

```
. lincom 1.female + 1.female#60.age3, irr
------------------------------------------------------------------------------
          _d |        IRR   Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         (1) |   .7065416   .0426289    -5.76   0.000     .6277414    .7952336
------------------------------------------------------------------------------

. lincom 1.female + 1.female#75.age3, irr
------------------------------------------------------------------------------
          _d |        IRR   Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         (1) |   .7206063   .0503884    -4.69   0.000     .6283155    .8264532
------------------------------------------------------------------------------
```

- The calculation $0.468 \times 2.098 = 0.707$ is not difficult but calculating the standard error and confidence intervals is non-trivial (a combination of variances and covariances). Lincom is useful for obtaining the CIs.

# Review: Parameterisations of the interaction model

- Interaction models can be parameterised in different ways to show effects using different reference groups.

- Interaction models measure deviations from the hypothesis of main effects, i.e. that the effect of sex is the same in all age categories.

- Interactions are symmetrical, meaning that we can choose either variable sex or age as the effect modifier of the other. The test for interaction will be the same regardless of parameterisation.

- If we wish to change reference group, simply divide the cells with corresponding reference group. We can choose any contrasts, i.e. compare any groups.

- For example: If we wish to tabulate the HRs, using the men aged 0-59y as reference group for all our HRs, then we simply divided all cells with the rate for men 0-59.

## Extra: How the two parameterisations are related

- This explains how the two parametrisations with ## and # for the age-specific hazard ratios for women vs men are mathematically related.

- Both parametrisations use 6 parameters, $\beta$ or $\gamma$, which have different interpretations, as they represent different contrasts (group comparisons).

| | female=0 (men) | female=1 (women) |
|---|---|---|
| 0-59 | $\exp(\beta_0) = \exp(\gamma_0)$ | $\exp(\beta_0 + \beta_1) = \exp(\gamma_0 + \gamma_3)$ |
| 60-74 | $\exp(\beta_0 + \beta_2) = \exp(\gamma_1)$ | $\exp(\beta_0 + \beta_1 + \beta_2 + \beta_4) = \exp(\gamma_0 + \gamma_4)$ |
| 75+ | $\exp(\beta_0 + \beta_3) = \exp(\gamma_2)$ | $\exp(\beta_0 + \beta_1 + \beta_3 + \beta_5) = \exp(\gamma_0 + \gamma_5)$ |

- The stratum-specific interaction model is $\ln(\lambda) = \gamma_0$age0 $+ \gamma_1$age60 $+$ $\gamma_2$age75 $+ \gamma_3$female x age0 $+ \gamma_4$female x age60 $+ \gamma_5$female x age75