

Lab 13*Poisson Regression --- examples*

These examples come from Chapter 12 of “Practical Biostatistical Methods” by Steve Selvin (1 edition (September 20, 1994)).

Datasets on the lecture website.

Example 12.11 Hodgkins disease mortality.

```
*****
* Prepare tbl12-11 for use in Stata
* copy/paste this dataset into the data editor

sex, pyears , deaths, age
1, 1299868, 55, 1
1, 1240595, 49, 2
1, 1045453, 38, 3
1, 795776, 26, 4
1, 645991, 19, 5
1, 599729, 17, 6
1, 568109, 22, 7
1, 506475, 21, 8
1, 368751, 18, 9
1, 252581, 11, 10
1, 140053, 10, 11
1, 81850, 4, 12
2, 1300402, 37, 1
2, 1217896, 29, 2
2, 1045801, 23, 3
2, 10260, 12, 4
2, 665612, 7, 5
2, 633646, 12, 6
2, 650686, 9, 7
2, 600455, 19, 8
2, 474609, 13, 9
2, 376781, 14, 10
2, 255412, 5, 11
2, 313603, 3, 12

* commands to set up dataset

label data "Hodgkins disease mortality data."

label variable sex "1=male, 2=female"
label define sexlab 1 male 2 female
label values sex sexlab

label variable pyears "Person-years (exposure)"

label variable deaths "Deaths (events)"

label variable age "Age groups 1=30-34, ..., 11=80-84, 12=85+ years"
label define agelab 1 "30-34 years" 2 "35-39 years" 3 "40-44 years" 4 "45-49 years"
5 "50-54 years" 6 "55-59 years" 7 "60-64 years" 8 "65-69 years" 9 "70-74 years" 10
"75-79 years" 11 "80-84 years" 12 "85+ years"
label values age agelab
```

* Poisson regression to model deaths as a function of sex and age,
 * with person-years (pyears) the exposure.

xi:poisson deaths i.sex i.age, exposure(pyears)

```
. xi:poisson deaths i.sex i.age, exposure(pyears)
i.sex      _Isex_1-2      (naturally coded; _Isex_1 omitted)
i.age      _Iage_1-12     (naturally coded; _Iage_1 omitted)
```

```
Iteration 0:  log likelihood = -101.32018
Iteration 1:  log likelihood = -94.501083
Iteration 2:  log likelihood = -94.436329
Iteration 3:  log likelihood = -94.436287
Iteration 4:  log likelihood = -94.436287
```

```
Poisson regression                                Number of obs   =          24
                                                  LR chi2(12)     =          43.98
                                                  Prob > chi2     =          0.0000
Log likelihood = -94.436287                    Pseudo R2      =          0.1889
```

deaths	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
_Isex_2 	-.4347787	.0975538	-4.46	0.000	-.6259806 - .2435768
_Iage_2	-.1110303	.1539164	-0.72	0.471	-.412701 .1906404
_Iage_3	-.1930717	.1651151	-1.17	0.242	-.5166914 .1305481
_Iage_4	.0975499	.196455	0.50	0.620	-.2874947 .4825946
_Iage_5 	-.5761639	.2221071	-2.59	0.009	-1.011486 - .1408419
_Iage_6 	-.4027727	.2129647	-1.89	0.059	-.8201758 .0146304
_Iage_7	-.3154855	.2076964	-1.52	0.129	-.722563 .091592
_Iage_8	.0394102	.1894352	0.21	0.835	-.3318759 .4106964
_Iage_9	.0653635	.2077584	0.31	0.753	-.3418355 .4725625
_Iage_10	.1588621	.2257462	0.70	0.482	-.2835923 .6013165
_Iage_11	.1339931	.2788291	0.48	0.631	-.4125018 .680488
_Iage_12	-.5585607	.3933177	-1.42	0.156	-1.329449 .2123278
_cons	-10.05535	.1110857	-90.52	0.000	-10.27307 -9.837625
pyears	(exposure)				

estat gof

```
Goodness-of-fit chi2 = 78.65163
Prob > chi2(11)     = 0.0000
```

Sex is significant, as well as age groups 5 and 6 (folks in their 50s).
 The goodness of fit test indicates the model does not fit well.
 Let's include the interaction.

* Poisson regression to model deaths as a function of sex, age, and their interaction,
 * with person-years (pyears) the exposure.

xi:poisson deaths i.sex i.age i.sex*i.age, exposure(pyears)

```
i.sex          _Isex_1-2          (naturally coded; _Isex_1 omitted)
i.age          _Iage_1-12         (naturally coded; _Iage_1 omitted)
i.sex*i.age    _IsexXage_#_#      (coded as above)
note: _Isex_2 dropped due to collinearity
note: _Iage_2 dropped due to collinearity
...
note: _Iage_11 dropped due to collinearity
note: _Iage_12 dropped due to collinearity
```

```
Iteration 0:  log likelihood = -55.700145
Iteration 1:  log likelihood = -55.11208
Iteration 2:  log likelihood = -55.110536
Iteration 3:  log likelihood = -55.110536
```

```
Poisson regression                                Number of obs   =          24
                                                  LR chi2(23)    =        122.64
                                                  Prob > chi2    =         0.0000
Log likelihood = -55.110536                      Pseudo R2      =         0.5267
```

deaths	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
_Isex_2 	-.396826	.2126237	-1.87	0.062	-.8135608 .0199088
_Iage_2	-.0688413	.1964433	-0.35	0.726	-.4538631 .3161806
_Iage_3	-.1519346	.2109446	-0.72	0.471	-.5653783 .2615092
_Iage_4	-.2585364	.2379986	-1.09	0.277	-.7250052 .2079324
_Iage_5	-.3636618	.2661079	-1.37	0.172	-.8852236 .1579001
_Iage_6	-.4005797	.2774984	-1.44	0.149	-.9444665 .1433071
_Iage_7	-.088586	.2522625	-0.35	0.725	-.5830114 .4058394
_Iage_8	-.0202677	.2565168	-0.08	0.937	-.5230314 .482496
_Iage_9	.142935	.2715463	0.53	0.599	-.3892859 .6751559
_Iage_10	.0288481	.3302891	0.09	0.930	-.6185067 .6762029
_Iage_11	.523249	.3437758	1.52	0.128	-.1505393 1.197037
_Iage_12	.1440909	.5178627	0.28	0.781	-.8709014 1.159083
_IsexXage~2	-.1092321	.3163855	-0.35	0.730	-.7293363 .510872
_IsexXage~3	-.1055988	.3391208	-0.31	0.756	-.7702633 .5590657
_IsexXage~4 	3.974701	.4086609	9.73	0.000	3.17374 4.775662
_IsexXage~5	-.6316242	.4906094	-1.29	0.198	-1.593201 .3299525
_IsexXage~6	-.0064932	.4328576	-0.02	0.988	-.8548786 .8418921
_IsexXage~7	-.6327058	.4491932	-1.41	0.159	-1.513108 .2476967
_IsexXage~8	.1265298	.3813915	0.33	0.740	-.6209838 .8740434
_IsexXage~9	-.1809661	.4215299	-0.43	0.668	-1.00715 .6452174
_IsexXage~10	.2380559	.4555727	0.52	0.601	-.6548502 1.130962
_IsexXage~11	-.8971782	.5875448	-1.53	0.127	-2.048745 .2543884
_IsexXage~12	-1.234096	.7928065	-1.56	0.120	-2.787968 .3197766
_cons	-10.07044	.13484	-74.68	0.000	-10.33472 -9.806159
pyears	(exposure)				

estat gof

```
Goodness-of-fit chi2 = .0001246
Prob > chi2(0)      = .
```

Sex is a little less important in the interaction model.

What really sticks out is females in age group 4 (45-49).

There are not enough degrees of freedom left to assess goodness-of-fit, since all 23 degrees of freedom were used to fit the model.

Let's look at interaction plots of this.

Keep in mind that we are modelling incident rates – not death counts (although related by the person-years). *We can't see what the model indicates if we look at the counts...*

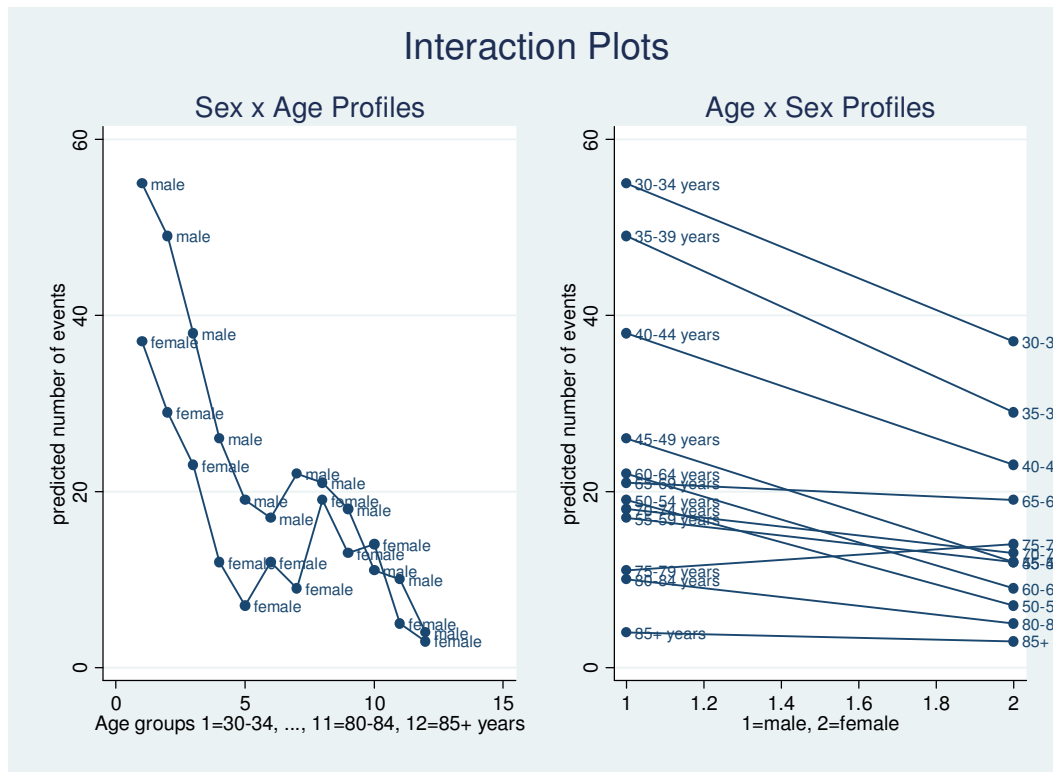
```
* these fitn values will be the same as the deaths
* since the interactions and main effects explicitly predict the original values
predict fitn, n

* interaction plot, sex is lines, age is horizontal axis
sort sex age
scatter fitn age, connect(ascending) mlabel(sex) title(Sex x Age Profiles)
scatter fitn age, connect(ascending) mlabel(sex) name(iplot1,replace) nodraw title(Sex x
Age Profiles)

* interaction plot, age is lines, sex is horizontal axis
sort age sex
scatter fitn sex, connect(ascending) mlabel(age) title(Age x Sex Profiles)
scatter fitn sex, connect(ascending) mlabel(age) name(iplot2,replace) nodraw title(Age x
Sex Profiles)

* combine the two interaction plots
graph combine iplot1 iplot2, title(Interaction Plots)
```

Death counts:



Is there interaction in the counts? It is impossible to tell since these plots do not account for the exposure rates (person-years). These plots do not correspond to the model, since the model is modelling death rates, not death counts. The next page gives the plots in incidence rate, $\exp(xb)$.

We are modelling incidence rates, $\exp(xb)$, so these are the interaction plots corresponding to our model.

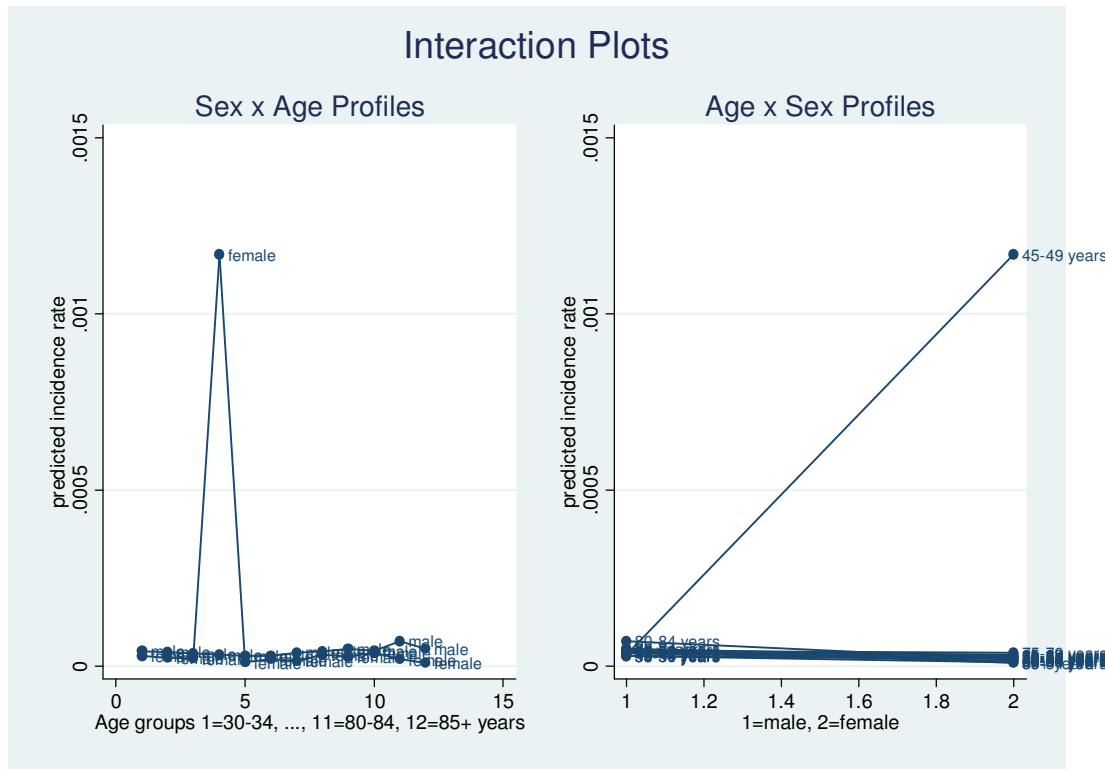
```
* predict the incident rates from the model
predict ir, ir

* interaction plot, sex is lines, age is horizontal axis
sort sex age
scatter ir age, connect(ascending) mlabel(sex) title(Sex x Age Profiles)
scatter ir age, connect(ascending) mlabel(sex) name(iplot1,replace) nodraw title(Sex x
Age Profiles)

* interaction plot, age is lines, sex is horizontal axis
sort age sex
scatter ir sex, connect(ascending) mlabel(age) title(Age x Sex Profiles)
scatter ir sex, connect(ascending) mlabel(age) name(iplot2,replace) nodraw title(Age x
Sex Profiles)

* combine the two interaction plots
graph combine iplot1 iplot2, title(Interaction Plots)
```

Incidence rates:



Well, nice to meet you females age group 4 (45-49). This corresponds to the model – a single huge interaction.

* Prepare tbl12-13 for use in Stata

* copy/paste this dataset into the data editor

```
age,msv,deaths,pyears
1, 1, 0, 29901
2, 1, 1, 6251
3, 1, 4, 5251
4, 1, 3, 4126
5, 1, 3, 2778
6, 1, 1, 1607
7, 1, 3, 1358
1, 2, 1, 71382
2, 2, 5, 16705
3, 2, 4, 13752
4, 2, 10, 10439
5, 2, 11, 7131
6, 2, 16, 4133
7, 2, 11, 3814
1, 3, 0, 6523
2, 3, 2, 2423
3, 3, 1, 2281
4, 3, 2, 1918
5, 3, 0, 1322
6, 3, 2, 723
7, 3, 3, 538
1, 4, 0, 2341
2, 4, 0, 972
3, 4, 1, 958
4, 4, 1, 816
5, 4, 0, 578
6, 4, 2, 375
7, 4, 3, 303
1, 5, 0, 1363
2, 5, 0, 478
3, 5, 0, 476
4, 5, 0, 387
5, 5, 0, 225
6, 5, 1, 164
7, 5, 0, 150
1, 6, 0, 779
2, 6, 0, 296
3, 6, 0, 282
4, 6, 1, 251
5, 6, 0, 193
6, 6, 0, 125
7, 6, 0, 69
1, 7, 0, 520
2, 7, 0, 188
3, 7, 0, 217
4, 7, 0, 184
5, 7, 1, 109
6, 7, 0, 60
7, 7, 1, 23
1, 8, 0, 2104
2, 8, 0, 1027
3, 8, 1, 1029
4, 8, 3, 827
5, 8, 1, 555
6, 8, 2, 297
7, 8, 2, 153
```

* commands to set up dataset

```
label data "Oak Ridge National Laboratory mortality."
```

```
label variable age "Age groups 1=<45, 2=45-49, ..., 6=65-69, 7=70+ years"
```

```
label define agelab 1 "<45 years" 2 "45-49 years" 3 "50-54 years" 4 "55-59 years" 5  
"60-64 years" 6 "65-69 years" 7 "70+ years"
```

```
label values age agelab
```

```
label variable msv "Exposure in mSv=milliseiverts (.1 rem), radiation"
```

```
label define msvlab 1 "mSv=0" 2 "mSv=0-19" 3 "mSv=20-39" 4 "mSv=40-59" 5 "mSv=60-79" 6  
"mSv=80-99" 7 "mSv=100-119" 8 "mSv=120+"
```

```
label values msv msvlab
```

```
label variable pyears "Person-years (exposure)"
```

```
label variable deaths "Deaths (events)"
```


* Poisson regression to model deaths as a function of exposure, age,
 * and their interaction, with person-years (pyears) the exposure.
xi:poisson deaths i.msv i.age i.msv*i.age, exposure(pyears)

```
i.msv          _Imsv_1-8          (naturally coded; _Imsv_1 omitted)
i.age          _Iage_1-7          (naturally coded; _Iage_1 omitted)
i.msv*i.age    _ImsvXage_#_#     (coded as above)
note: _Imsv_2 dropped due to collinearity
...
note: _Iage_7 dropped due to collinearity
```

```
Iteration 0:  log likelihood = -241.44813
Iteration 1:  log likelihood = -166.32715
...
Iteration 9:  log likelihood = -42.462632
Iteration 10: log likelihood = -42.46263
```

```
Poisson regression          Number of obs   =          56
LR chi2(55)                 =          269.42
Prob > chi2                 =           0.0000
Pseudo R2                   =           0.7603

Log likelihood = -42.46263
```

deaths	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
_Imsv_2	16.74431	6682.334	0.00	0.998	-13080.39 13113.88
_Imsv_3	1.522558	9450.245	0.00	1.000	-18520.62 18523.66
_Imsv_4	2.547315	9450.244	0.00	1.000	-18519.59 18524.69
_Imsv_5	3.088205	9450.244	0.00	1.000	-18519.05 18525.23
_Imsv_6	3.647637	9450.244	0.00	1.000	-18518.49 18525.79
_Imsv_7	4.051819	9450.244	0.00	1.000	-18518.09 18526.19
_Imsv_8	2.654053	9450.244	0.00	1.000	-18519.48 18524.79
_Iage_2	19.17961	6682.334	0.00	0.998	-13077.95 13116.31
_Iage_3	20.74022	6682.333	0.00	0.998	-13076.39 13117.87
_Iage_4	20.69366	6682.333	0.00	0.998	-13076.44 13117.83
_Iage_5	21.08924	6682.333	0.00	0.997	-13076.04 13118.22
_Iage_6	20.53799	6682.334	0.00	0.998	-13076.6 13117.67
_Iage_7	21.80495	6682.333	0.00	0.997	-13075.33 13118.94
_ImsvXage~2_2	-16.11784	6682.334	-0.00	0.998	-13113.25 13081.02
_ImsvXage~2_3	-17.70707	6682.334	-0.00	0.998	-13114.84 13079.43
_ImsvXage~2_4	-16.4685	6682.334	-0.00	0.998	-13113.6 13080.66
_ImsvXage~2_5	-16.38775	6682.334	-0.00	0.998	-13113.52 13080.75
_ImsvXage~2_6	-14.91636	6682.334	-0.00	0.998	-13112.05 13082.22
_ImsvXage~2_7	-16.4777	6682.334	-0.00	0.998	-13113.61 13080.66
_ImsvXage~3_2	.118307	9450.245	0.00	1.000	-18522.02 18522.26
_ImsvXage~3_3	-2.075034	9450.245	-0.00	1.000	-18524.21 18520.06
_ImsvXage~3_4	-1.162016	9450.245	-0.00	1.000	-18523.3 18520.98
_ImsvXage~3_5	-19.49305	11574.14	-0.00	0.999	-22704.39 22665.4
_ImsvXage~3_6	-.0307128	9450.245	-0.00	1.000	-18522.17 18522.11
_ImsvXage~3_7	-.5966487	9450.245	-0.00	1.000	-18522.74 18521.54
_ImsvXage~4_2	-18.30064	11574.14	-0.00	0.999	-22703.19 22666.59
_ImsvXage~4_3	-2.232269	9450.245	-0.00	1.000	-18524.37 18519.91
_ImsvXage~4_4	-2.025279	9450.245	-0.00	1.000	-18524.16 18520.11
_ImsvXage~4_5	-19.69048	11574.14	-0.00	0.999	-22704.58 22665.2
_ImsvXage~4_6	-.3989858	9450.245	-0.00	1.000	-18522.54 18521.74
_ImsvXage~4_7	-1.047279	9450.244	-0.00	1.000	-18523.19 18521.09
_ImsvXage~5_2	-18.13178	11574.14	-0.00	0.999	-22703.02 22666.76
_ImsvXage~5_3	-19.68819	11574.14	-0.00	0.999	-22704.58 22665.2
_ImsvXage~5_4	-19.43464	11574.14	-0.00	0.999	-22704.33 22665.46
_ImsvXage~5_5	-19.28789	11574.14	-0.00	0.999	-22704.18 22665.6
_ImsvXage~5_6	-.8059468	9450.245	-0.00	1.000	-18522.94 18521.33
_ImsvXage~5_7	-19.59815	11574.14	-0.00	0.999	-22704.49 22665.29
_ImsvXage~6_2	-18.21196	11574.14	-0.00	0.999	-22703.1 22666.68
_ImsvXage~6_3	-19.72411	11574.14	-0.00	0.999	-22704.62 22665.17
_ImsvXage~6_4	-1.94664	9450.244	-0.00	1.000	-18524.09 18520.19
_ImsvXage~6_5	-19.69392	11574.14	-0.00	0.999	-22704.59 22665.2
_ImsvXage~6_6	-18.70829	11574.14	-0.00	0.999	-22703.6 22666.18
_ImsvXage~6_7	-19.38105	11574.14	-0.00	0.999	-22704.27 22665.51
_ImsvXage~7_2	-18.16223	11574.14	-0.00	0.999	-22703.05 22666.73
_ImsvXage~7_3	-19.86628	11574.14	-0.00	0.999	-22704.76 22665.03
_ImsvXage~7_4	-19.65477	11574.14	-0.00	0.999	-22704.55 22665.24
_ImsvXage~7_5	-1.912294	9450.245	-0.00	1.000	-18524.05 18520.23
_ImsvXage~7_6	-18.3785	11574.14	-0.00	0.999	-22703.27 22666.51
_ImsvXage~7_7	-1.072159	9450.245	-0.00	1.000	-18523.21 18521.07
_ImsvXage~8_2	-18.46241	11574.14	-0.00	0.999	-22703.36 22666.43
_ImsvXage~8_3	-2.410501	9450.245	-0.00	1.000	-18524.55 18519.73
_ImsvXage~8_4	-1.046794	9450.244	-0.00	1.000	-18523.19 18521.09
_ImsvXage~8_5	-2.142148	9450.245	-0.00	1.000	-18524.28 18520
_ImsvXage~8_6	-.2725299	9450.245	-0.00	1.000	-18522.41 18521.87
_ImsvXage~8_7	-.8762051	9450.244	-0.00	1.000	-18523.02 18521.26
_cons	-27.92009	6682.333	-0.00	0.997	-13125.05 13069.21
pyears	(exposure)				

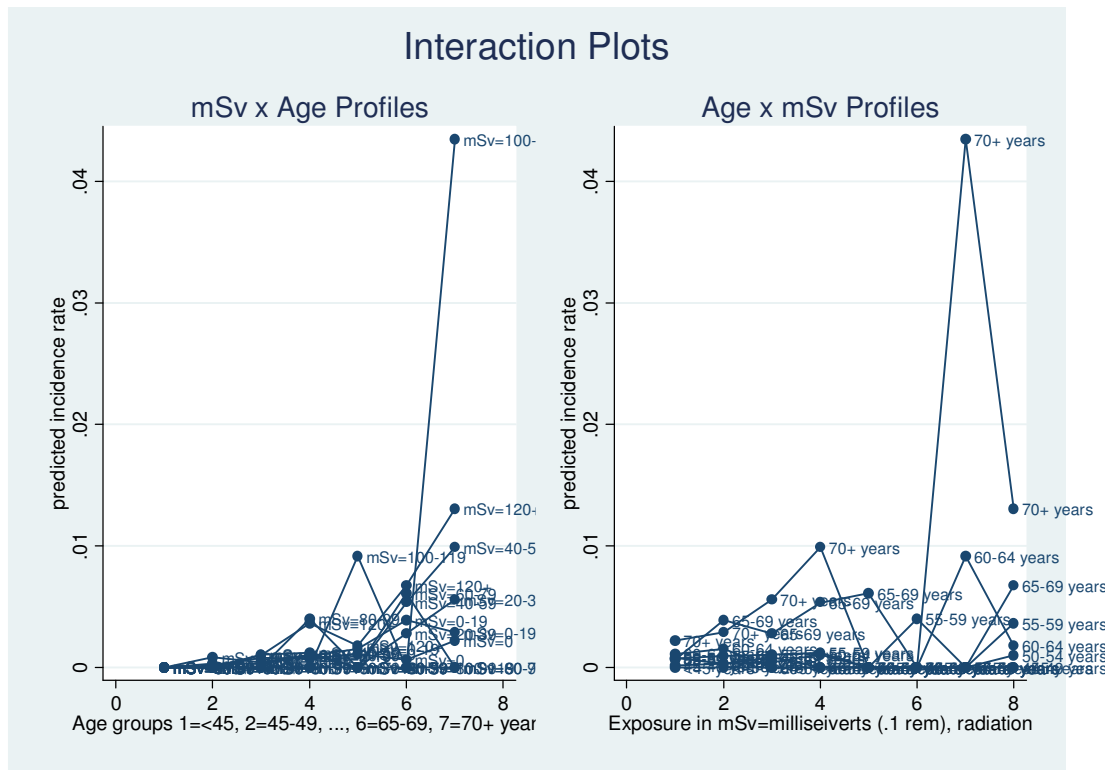
The interaction model has extremely large standard errors, therefore nothing appears significant. However, when we look at the interaction plots, we see what is going on.

```
* predict the incident rates from the model
predict ir, ir

* interaction plot, msv is lines, age is horizontal axis
sort msv age
scatter ir age, connect(ascending) mlabel(msv) title(mSv x Age Profiles)
scatter ir age, connect(ascending) mlabel(msv) name(iplot1,replace) nodraw title(mSv x
Age Profiles)

* interaction plot, age is lines, msv is horizontal axis
sort age msv
scatter ir msv, connect(ascending) mlabel(age) title(Age x mSv Profiles)
scatter ir msv, connect(ascending) mlabel(age) name(iplot2,replace) nodraw title(Age x
mSv Profiles)

* combine the two interaction plots
graph combine iplot1 iplot2, title(Interaction Plots)
```



The incidence rate is extremely high for age group 8 (70+ years) those with exposure level 7 (mSv=100-119). The left plot also indicates that the incidence rate tends to increase as age increases. That can also be seen in the right plot since those lines that rise from the bottom belong to the older age groups.