

# Models In Epidemiology And Biostatistics

## Gordon Hilton Fick

### Exposure with More Than 2 Levels

Lets begin this illustration by considering D: stroke (Y or N) and E: stenosis=stengrp (1, 2, 3) with age group and gender as potential confounders/modifiers. (This is only the start)

```
. egen genage = group(sex agegp)
. label def gal 1 "YF" 2 "OF" 3 "VOF" 4 "YM" 5 "OM" 6 "VOM"
. label val genage gal
```

A 'classical' stratified analysis could begin by creating the estimates of the rate ratios, considering the levels of exposure 2 at a time [1 cf 2 ignoring 3; 1 cf 3 ignoring 2 and 2 cf 3 ignoring 1]

```
. gen st12=stengrp
. recode st12 1=0 2=1 3=.
(st12: 724 changes made)
. gen st13=stengrp
. recode st13 1=0 2=. 3=1
(st13: 724 changes made)
. gen st23=stengrp
. recode st23 1=. 2=0 3=1
(st23: 724 changes made)
. cs stro st12,by(genage)
```

group(sex agegp)	RR	[95% Conf. Interval]		M-H Weight
YF	.3015873	.0400535	2.270838	2.25
OF	1.573333	.4856241	5.097312	1.785714
VOF	.5333333	.063967	4.446741	1.153846
YM	1.409091	.6208763	3.197959	3.854015
OM	2.545455	1.332577	4.862261	4.888889
VOM	1.623932	.7672776	3.437027	3.65625
Crude	1.634694	1.133961	2.35654	
M-H combined	1.587159	1.093676	2.303308	
Test of homogeneity (M-H)      chi2(5) =      5.746    Pr>chi2 = 0.3317				

```
. cs stro st13,by(genage)
```

group(sex agegp)	RR	[95% Conf. Interval]		M-H Weight
YF	.5816327	.1647614	2.053251	2.969697
OF	0	.	.	1.461538
VOF	2.37037	.67603	8.311252	1.08
YM	1.519608	.7060116	3.270779	4.25
OM	2.444444	1.169454	5.109486	3.272727
VOM	1.333333	.5563967	3.195162	3
Crude	1.47036	.9884287	2.187267	
M-H combined	1.418589	.9577273	2.10122	
Test of homogeneity (M-H)      chi2(5) =      4.705    Pr>chi2 = 0.4529				

```
. cs stro st23,by(genage)
```

group(sex agegp)	RR	[95% Conf. Interval]		M-H Weight
YF	1.928571	.2170648	17.13492	.6086957
OF	0	.	.	1.727273
VOF	4.444444	.6028686	32.76516	.4736842
YM	1.078431	.4666538	2.492242	4.294737
OM	.9603175	.5139499	1.794357	7.411765
VOM	.8210526	.3611643	1.86654	4.222222
Crude	.8994709	.6057626	1.335586	
M-H combined	.9870156	.6612479	1.473275	
Test of homogeneity (M-H)      chi2(5) =      2.784    Pr>chi2 = 0.7332				

It appears that the rate ratios based on a comparison of either stengrp==2 with stengrp==1 or stengrp==3 with stengrp==1 show that rate ratio may be meaningfully greater than one for the middle age group males while for each of the other 5 groups, the rate ratio might well be one. It maybe worth considering the combining of stengrp==2 with stengrp==3 [and hence return to the “simpler” 2 exposure levels scenario]

```
. tab stro stengrp if genage==5,col
```

stro	STENGRP			Total
	1	2	3	
0	100	56	31	187
	89.29	72.73	73.81	80.95
1	12	21	11	44
	10.71	27.27	26.19	19.05
Total	112	77	42	231
	100.00	100.00	100.00	100.00

```
. disp 27.27/10.71
2.5462185
```

```
. disp 26.19/10.71
2.4453782
```

Now lets see if such observations are maintained if we were to use 'odds ratios' instead of 'rate ratios'

```
. cc stro st12,by(genage)
```

group(sex agegp)	OR	[95% Conf. Interval]		M-H Weight
YF	.2605042	.0054641	2.359816	2.125 (exact)
OF	1.68254	.3141825	7.89961	1.5 (exact)
VOF	.4814815	.0082672	7.367561	1.038462 (exact)
YM	1.5	.4856993	4.38596	3.153285 (exact)
OM	3.125	1.343246	7.481712	3.555556 (exact)
VOM	2.013889	.589832	6.87899	2.25 (exact)
Crude	1.81896	1.130739	2.917292	(exact)
M-H combined	1.758122	1.117418	2.766193	
Test of homogeneity (M-H)      chi2(5) =      6.31    Pr>chi2 = 0.2776				

Test that combined OR = 1:

```
      Mantel-Haenszel chi2(1) =      5.97
                         Pr>chi2 =      0.0145
```

```
. cc stro st13,by(genage)
```

group(sex agegp)	OR	[95% Conf. Interval]		M-H Weight	
YF	.5314286	.0811707	2.651915	2.651515	(exact)
OF	0	0	1.897777	1.461538	(exact)
VOF	3.466667	.3970454	31.55774	.6	(exact)
YM	1.646341	.5819084	4.537637	3.416667	(exact)
OM	2.956989	1.059069	8.088757	2.415584	(exact)
VOM	1.487179	.3540552	5.86488	2.052632	(exact)
Crude	1.589705	.948017	2.640215		(exact)
M-H combined	1.532757	.9423053	2.493189		
-----					
Test of homogeneity (B-D)		chi2(5) =	8.17	Pr>chi2 =	0.1473

```
. cc stro st23,by(genage)
```

group(sex agegp)	OR	[95% Conf. Interval]		M-H Weight	
YF	2.04	.1469205	113.2837	.5434783	(exact)
OF	0	0	1.163063	1.727273	(exact)
VOF	7.2	.4685623	392.9746	.2631579	(exact)
YM	1.097561	.3464983	3.574656	3.452632	(exact)
OM	.9462366	.362119	2.37916	5.470588	(exact)
VOM	.7384615	.1720874	3.014251	2.888889	(exact)
Crude	.8739635	.5112596	1.486066		(exact)
M-H combined	.9830401	.5845476	1.653189		
-----					
Test of homogeneity (B-D)		chi2(5) =	6.89	Pr>chi2 =	0.2292

It would seem that the message is the same with odds ratios. Here is an 'incorrect' logistic regression version of the above. Note the 'note' and the deletion.

```
. xi:logit stro i.stengrp*i.genage
i.stengrp      _Istengrp_1-3      (naturally coded; _Istengrp_1 omitted)
i.genage        _Igenage_1-6      (naturally coded; _Igenage_1 omitted)
i.ste~p*i.gen~e  _IsteXgen_#_#    (coded as above)
```

note: \_IsteXgen\_3\_2 != 0 predicts failure perfectly  
\_IsteXgen\_3\_2 dropped and 19 obs not used

Logistic regression	Number of obs	=	705
	LR chi2(16)	=	31.31
	Prob > chi2	=	0.0123
Log likelihood = -318.34029	Pseudo R2	=	0.0469

stro	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_Istengrp_2	-1.345136	1.110826	-1.21	0.226	-3.522315	.8320428
_Istengrp_3	-.6321865	.7405731	-0.85	0.393	-2.083683	.8193102
_Igenage_2	-.6904554	.6005413	-1.15	0.250	-1.867495	.4865838
_Igenage_3	.02174	.7650958	0.03	0.977	-1.47782	1.5213
_Igenage_4	-.4214654	.5203789	-0.81	0.418	-1.441389	.5984585
_Igenage_5	-.6321865	.5181202	-1.22	0.222	-1.647683	.3833104
_Igenage_6	.3180058	.5663118	0.56	0.574	-.791945	1.427957
_IsteXge~2_2	1.865441	1.310377	1.42	0.155	-.7028513	4.433733
_IsteXge~2_3	.6142488	1.65991	0.37	0.711	-2.639116	3.867613
_IsteXge~2_4	1.750601	1.217535	1.44	0.150	-.6357233	4.136926
_IsteXge~2_5	2.484571	1.180146	2.11	0.035	.1715273	4.797614
_IsteXge~2_6	2.045204	1.241784	1.65	0.100	-.3886488	4.479057
_IsteXge~3_3	1.87538	1.186889	1.58	0.114	-.4508792	4.201639
_IsteXge~3_4	1.130742	.8766515	1.29	0.197	-.5874633	2.848947
_IsteXge~3_5	1.716358	.8746136	1.96	0.050	.0021468	3.430569
_IsteXge~3_6	1.029068	.9683141	1.06	0.288	-.8687929	2.926929
_cons	-1.488077	.4184677	-3.56	0.000	-2.308259	-.6678955

Lets try the grouping of stengroups

```
. gen stgt1=(stengrp>1)
```

```
. cc stro stgt1,by(genage)
```

group(sex agegp)	OR	[95% Conf. Interval]		M-H Weight
YF	.4217687	.0838541	1.850763	3.5 (exact)
OF	.8833333	.1717266	4.017909	2.330097 (exact)
VOF	1.547619	.2392606	11.86432	1.2 (exact)
YM	1.577922	.6666959	3.839175	4.914894 (exact)
OM	3.065134	1.424533	6.925176	4.519481 (exact)
VOM	1.777778	.615227	5.325877	3.144578 (exact)
Crude	1.712662	1.139264	2.588672	(exact)
M-H combined	1.661992	1.122119	2.461607	

Test of homogeneity (M-H)      chi2(5) =      7.85    Pr>chi2 = 0.1646

Test that combined OR = 1:

Mantel-Haenszel    chi2(1) =      6.51  
Pr>chi2 =      0.0107

```
. xi:logit stro i.stgt1*i.genage
```

```
i.stgt1                _Istgt1_0-1                (naturally coded; _Istgt1_0 omitted)
i.genage                _Igenage_1-6                (naturally coded; _Igenage_1 omitted)
i.stgt1*i.gen~e        _IstgXgen_#_#                (coded as above)
```

Logistic regression	Number of obs	=	724
	LR chi2(11)	=	30.29
	Prob > chi2	=	0.0014
Log likelihood = -322.60145	Pseudo R2	=	0.0448

stro	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_Istgt1_1	-.8632982	.6700185	-1.29	0.198	-2.17651	.4499138
_Igenage_2	-.6904554	.6005413	-1.15	0.250	-1.867495	.4865838
_Igenage_3	.02174	.7650958	0.03	0.977	-1.47782	1.5213
_Igenage_4	-.4214654	.5203789	-0.81	0.418	-1.441389	.5984585
_Igenage_5	-.6321865	.5181202	-1.22	0.222	-1.647683	.3833104
_Igenage_6	.3180058	.5663118	0.56	0.574	-.791945	1.427957
_IstgXgen_~2	.7392456	.9536558	0.78	0.438	-1.129885	2.608377
_IstgXgen_~3	1.300016	1.063301	1.22	0.221	-.7840165	3.384048
_IstgXgen_~4	1.319407	.7830366	1.68	0.092	-.2153165	2.854131
<b>_IstgXgen_~5  </b>	<b>1.98339</b>	<b>.7648544</b>	<b>2.59</b>	<b>0.010</b>	<b>.4843024</b>	<b>3.482477</b>
_IstgXgen_~6	1.438662	.8315656	1.73	0.084	-.1911763	3.068501
_cons	-1.488077	.4184677	-3.56	0.000	-2.308259	-.6678955

```
. xi:logit stro i.stgt1*i.genage,or
i.stgt1      _Istgt1_0-1      (naturally coded; _Istgt1_0 omitted)
i.genage      _Igenage_1-6      (naturally coded; _Igenage_1 omitted)
i.stgt1*i.gen~e  _IstgXgen_#_#  (coded as above)
```

```
Logistic regression              Number of obs   =       724
                                LR chi2(11)      =       30.29
                                Prob > chi2       =       0.0014
Log likelihood = -322.60145      Pseudo R2    =       0.0448
```

stro	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
_Istgt1_1	.4217687	.2825928	-1.29	0.198	.1134367	1.568177
_Igenage_2	.5013477	.30108	-1.15	0.250	.1545103	1.626749
_Igenage_3	1.021978	.7819111	0.03	0.977	.2281344	4.578174
_Igenage_4	.6560847	.3414126	-0.81	0.418	.2365988	1.819312
_Igenage_5	.5314286	.2753439	-1.22	0.222	.1924953	1.467133
_Igenage_6	1.374384	.77833	0.56	0.574	.4529629	4.170169
_IstgXgen_~2	2.094355	1.997294	0.78	0.438	.3230703	13.57699
_IstgXgen_~3	3.669355	3.90163	1.22	0.221	.4565685	29.48991
_IstgXgen_~4	3.741202	2.929498	1.68	0.092	.8062862	17.35934
_IstgXgen_~5	7.267334	5.558453	2.59	0.010	1.623042	32.54021
_IstgXgen_~6	4.215054	3.505094	1.73	0.084	.825987	21.50964

```
. disp 7.267334*0.4217687
3.065134
```

which is the estimated OR for the middle aged males. Further we get a direct message ( $p=0.010$ ) that the OR for the middle aged males is different from the other 5 groups in so far as we are comparing the OM group to the YF group and none of the other OR comparisons with YF have low p-values.