PharmaSUG 2016 – QT22

Using PROC GENMOD with count data

Meera G Kumar, Sanofi, Bridgewater, NJ

ABSTRACT

Use of PROC GENMOD in clinical trials data is quite common and more straightforward due to the availability of patient level data. How do you use the procedure to calculate event rate ratio using counts data? The key is to set up dummy variables for each dose level along with the 'offset' option. There can be situations in the Epidemiology area where you get only summary data for the number of events in each dose group or treatment arm. In the example discussed here, there is the unique situation where the numerator is not necessarily part of the denominator. This happens because the event data is pooled from various sources. This paper demonstrates how to use the count data and set up a Poisson distribution for the calculation of rate ratio along with the confidence interval and its associated p-value.

INTRODUCTION

The data of adverse events for drugs are sometimes available only in aggregate form, and such data are often derived from different sources. In this example, the number of an adverse event associated with each dose of drug A was derived from an aggregate report of adverse events. The number of patients who have used the drug of each dose was derived from the sales data in the same time period for the aggregate report of adverse events.

OBJECTIVE

The objective of the analysis was to compare the rate of different adverse events in each dose group against a reference dose group. Calculations of event rate, event rate ratio, 95% confidence interval and p-value were the parameters of interest. Input data (for the numerator) with event counts were provided in a table (Table 1) as follows:

Table 1. Adverse event (AE) by SOC1 for Drug A by dosage

Preferred Term	Total event count	doseA	doseB	doseC	doseD
AE1	6	3	0	1	2
AE2	56	18	12	17	9
AE3	21	3	5	5	8
AE4	19	4	1	3	11
AE5	12	1	2	2	7

Input data (for the denominator) providing the number of patients and their associated person-years was also provided in a table (Table 2) as follows:

Table 2. Cumulative patient exposure for Drug A by dosage*

Dose level	Patients	PY**
doseA	17,500	9,100
doseB	21,300	19,800
doseC	15,500	14,600
doseD	17,300	18,800

* = Worldwide data

**PY=Person-years. It is the sum of individual units of time that the persons in the study population have been exposed or at risk to the conditions of interest.

Data from the above tables were read into SAS datasets and the event rate and event rate ratio were calculated as follows.

Event rate=Numerator data/Denominator data, expressed as events/100,000 persons or events/100,000 person-years.

The SAS code below only shows the relevant calculation syntax. All other data clean up, proc transpose, merges etc. are not shown.

```
  data abc;
       set temp;
       /* calculate event rate as number of persons for each dose group */
       patrateA=doseApats/totpatsA*100000;
       patrateB=doseBpats/totpatsB*100000;
       patrateC=doseCpats/totpatsC*100000;
       patrateD=doseDpats/totpatsD*100000;

       /* calculate event rate as number of person-years for each dose group */
       pyrateA=doseApats/totpyA*100000;
       pyrateB=doseBpats/totpyB*100000;
       pyrateC=doseCpats/totpyC*100000;
       pyrateD=doseDpats/totpyD*100000;
   run;
```

Once the total counts were calculated for each dose group, dummy exposure variable was created with a value of 1 or 0. For example, if the ratio of doseA vs. doseB was compared, then the dummy exposure variable was set to be 0 for the reference treatment group. Next, log of the total person counts or total person-years was created. This step is essential to use the offset statement.

```
   data xyz;
        set abc;
        if dosegrp='A' then dumyexp=1;
        else if dosegrp='B' then dumyexp=0;
         …
         …
         …
        logpatA=log(totpatsA);
        logpatB=log(totpatsB);
   run;
```

PROC GENMOD was used to calculate the event rate ratio and the 95% Poisson confidence interval along with the p-value.

The variable 'aecnt' in the model statement below refers to the event count from Table 1 above. Variable logpatcnt contains the value of the log of the total count.

```
ods listing close;
proc genmod data=compAtoB;
  model aecnt = dumyexp / dist=poisson link=log offset=logpatcnt;
  estimate "Ratio doseA vs. doseB" dumyexp 1/exp;
  output out=rrstats;
  ods output Estimates=stats (keep=label meanestimate meanlowercl meanuppercl
                              probchisq
                              rename=(meanestimate=rateratio meanlowercl=lowci
                              meanuppercl=upci probchisq=pval));
run;
ods listing;
```

The code above was repeated for the comparison of doseA vs. doseC and doseA vs. doseD. It was also repeated for the calculation of rate ratio using person-years data.

Sample output results from PROC GENMOD as below:

The GENMOD procedure

Model Information	
Data Set	WORK.COMPATOB
Distribution	Poisson
Link Function	Log
Dependent Variable	aecnt
Offset Variable	logpatcnt

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-15.8140	1.0000	-17.7740	-13.8540	250.08	<.0001
dumyexp	1	0.4312	1.2247	-1.9692	2.8317	0.12	0.5218
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

Contrast Estimate Results										
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
Ratio doseA vs.doseB	1.3293	0.1263	15.9318	0.4312	1.2247	0.05	-1.9692	2.8317	0.12	0.5218
Exp(Ratio doseA vs.doseB)				1.5391	1.8850	0.05	0.1396	16.9738		

The layout of the final results table was as follows:

Table 4.1 Event rate for SOC1 (expressed as number of persons)													Rate Ratio					
	DoseA			DoseB			DoseC			DoseD			DoseA vs. DoseB		DoseA vs. DoseC		DoseA vs. DoseD	
Event	No. of events	No. of patients	Rate (per 100,000)	No. of events	No. of patients	Rate (per 100,000)	No. of events	No. of patients	Rate (per 100,000)	No. of events	No. of patients	Rate (per 100,000)	Ratio (95% CI)	p-value	Ratio (95% CI)	p-value	Ratio (95% CI)	p-value
Table 4.2 Event rate for SOC1 (expressed as number of person-years)													Rate Ratio					
	DoseA			DoseB			DoseC			DoseD			DoseA vs. DoseB		DoseA vs. DoseC		DoseA vs. DoseD	
Event	No. of events	No. of PYs	Rate (per 100,000)	No. of events	No. of PYs	Rate (per 100,000)	No. of events	No. of PYs	Rate (per 100,000)	No. of events	No. of PYs	Rate (per 100,000)	Ratio (95% CI)	p-value	Ratio (95% CI)	p-value	Ratio (95% CI)	p-value

CONCLUSION

The key technique to the analysis of counts data is the setup of dummy exposure variables for each dose level compared along with the 'offset' option. As demonstrated in the paper, it is quite simple to use PROC GENMOD with counts data. The implementation is quite simple and straightforward once it is known.

ACKNOWLEDGMENTS

I would like to thank Dr. Chuntao Wu for his guidance, support and input in writing the section for background information.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Meera G Kumar
Enterprise: Sanofi
Address: 55 Corporate Drive
City, State ZIP: Bridgewater, NJ 08807
Work Phone: (908)981-5190
E-mail: meera.kumar@sanofi.com