1998 Joint Statistical Meetings (Dallas)

Power for Categorical Data Analysis:

A Simpler Method to Compute Power for Likelihood Ratio Tests in Generalized Linear Models

Ralph G. O'Brien Gwowen Shieh*

Department of Biostatistics and Epidemiology

THE CLEVELAND CLINIC FOUNDATION

*Work completed while on leave from Dept of Management Science, National Chiao Tung University, Hsinchu, Taiwan, ROC

Logistic Regression Example

(fictitious, but realistic)

- Monoclonal antibody MA710808
- Possible treatment for patients in critical condition with severe sepsis

Proposed Design

- Randomized, placebo controlled, blinded Phase III trial
- 2/3 will get MA710808; 1/3 will get placebo

Predictor Variables

- **SevIndx0.** Several baseline measures collected and combined to create a baseline severity index, SevIndx0. SevIndx0 is a continuous measure developed in the previous Phase II trial and in this population ranges from 1.5 (least severe) to 9.5 with a tri-modal shape. For the sake of pragmatism, we shall simplify somewhat and pretend that its distribution is discrete:

SevIndx0:	2	3	4	5	6	7	8	9
Prob:	.10	.10	.05	.25	.25	.05	.10	.10

- **MA710808.** Code the dummy variable:
  MA710808 = 0, if received placebo
  MA710808 = 1, if received MA710808

- Thus, Pr[SevIndx0 = 2, MA710808 = 0] =
  1/10 * 1/3 = 1/30 = 2/60.

Assumed Model for Sample-Size Analysis

- From the Phase II results we take the linear predictor to be

  logit = 2.5 - 0.5*SevIndx0 + ln(3)*MA710808.

  The specification "ln(3)*MA710808" sets an odds ratio of 3 in favor of MA710808. This is a large effect, but actually less than the odds ratio of almost 5.0 observed in the Phase II study.

Steps to Perform Complex *yet Practical* Power Analyses

1. Formulate the problem completely. This is the hardest part.
2. Create exemplary data set for the problem. Straightforward programming.
3. "Analyze" exemplary data set using ordinary software. A dry run of the planned data analysis.
4. Convert test statistics to noncentrality values and then to powers (for given Ns) or sufficient Ns (for given powers). Easy with UnifyPow, a freeware SAS module/macro. (http://www.bio.ri.ccf.org/power.html)

2. Create exemplary data set for the problem.

- **Exemplary data set:** one that conforms exactly to the assumed model. Its "sample" estimates are identical to the population parameters.

- Via the SAS System.

```
data sepsis;
keep SevIndx0 MA710808 probX Alive25 probY ExmpCnt;
input SevIndx0 ProbSevI;
TotNExmp = 10002;  *Total N of exemplary data;
RanRateP = 1/3;  *Randomization Rate for the Placebo group;
betaInt = 2.5;
betaSev = -0.5;
betaMA = log(3);  *Note: ln(3) = 1.0986;
/*
Each dataline gives one SevIndx0 group. From each we generate
"exemplary" counts, ExmpCnt, for 4 types of cases based on the total
exemplary sample size, TotNExmp:
    (1) Cases on placebo who are not alive at Day 25.
    (2) Cases on placebo who are alive at Day 25.
    (3) Cases on MA710808 who are not alive at Day 25.
    (4) Cases on MA710808 who are alive at Day 25.
*/
do MA710808 = 0 to 1;  * 1 = randomized to MA710808;
  logit = betaInt + betaSev*SevIndx0 + betaMA*MA710808;
  PrAliv25 = 1/(1 + exp(-logit));  *Prob that case is alive at Day 25;
  if MA710808 = 0 then RandRate = RanRateP;
  else RandRate = 1 - RanRateP;
  probX = RandRate*ProbSevI;
  do Alive25 = 0 to 1;  * 1 = alive at Day 25;
    if Alive25 = 0 then probY = (1-PrAliv25);
    else probY = PrAliv25;
    ExmpCnt = round(TotNExmp*probX*probY);  *Exemplary count;
    output;
    end;
  end;
```

Do not bother with during presentation!


```
*SevIndx0  ProbSevI  ;     datalines;
     2       .10
     3       .10
     4       .05
     5       .25
     6       .25
     7       .05
     8       .10
     9       .10
```

- **The data set created.** 32 prototypical cases comprise the exemplary data set. The variable ExmpCnt gives the exemplary counts based on N_{total} = 10,002 (= 10,005 due to rounding).

Case	SevIndx0	MA710808	ProbX	Alive25	ProbY	ExmpCnt
1	2	0	0.03333	0	0.18243	61
2	2	0	0.03333	1	0.81757	273
3	2	1	0.06667	0	0.06923	46
4	2	1	0.06667	1	0.93077	621
5	3	0	0.03333	0	0.26894	90
<<<<<< cases 6-30 not shown >>>>>>						
29	9	0	0.03333	0	0.88080	294
30	9	0	0.03333	1	0.11920	40
31	9	1	0.06667	0	0.71123	474
32	9	1	0.06667	1	0.28877	193

3. "Analyze" exemplary data set using ordinary software.

- Input (SAS):

```
proc genmod;
  model Alive25 = SevIndx0 MA710808 /
        dist = binomial
        link = logit
        type3;
  freq ExmpCnt;
```


- Key lines from output:

```
Sum Of Frequency Weights          10005
```

LR Statistics For Type 3 Analysis

Source	DF	ChiSquare	Pr>Chi
SEVINDX0	1	1820.8867	0.0001
MA710808	1	528.9458	0.0001

4. Convert test statistics to powers or sufficient Ns.

```
title2 "Sample-size analysis for MA710808
Phase III study";
%include "&UnifyPow";
datalines;
Exemplary chi**2
Nexemplary 10005 . Not 10002 due to rounding.
alpha .01 .05
power .80 .90 .95
effects
"MA710808 vs. Placebo"  1   528.9458
%tables;
```


Scenario: Exemplary chi**2

		ALPHA					
		0.01			0.05		
		Minimum Power			Minimum Power		
		.800	.900	.950	.800	.900	.950
		Total N	Total N	Total N	Total N	Total N	Total N
Effect	Statistic						
MA710808 vs. Placebo	2-tail Z	222	282	338	149	199	247
	1-tail Z	190	247	299	117	162	205

Note that these are not necessarily multiples of 3. Thus they do not conform to 1:2 randomization.

```
%include "&UnifyPow";
title3 "Get powers for NTotals that fit 1:2
randomization";
datalines;
Exemplary chi**2
Nexemplary 10005
alpha .01
NTotal 222 282 339 . multiples of 3.
effects
"MA710808 vs. Placebo"  1   528.9458
%tables;
```


Scenario: Exemplary chi**2

		ALPHA		
		0.01		
		Total N		
		222	282	339
		Power	Power	Power
Effect	Statistic			
MA710808 vs. Placebo	2-tail Z	.802	.901	.951
	1-tail Z	.864	.938	.972

Poisson Regression Example

(fictitious, but realistic)

- **8 large identical computers.** The Reliable Web Server Company is relying on 8 large identical computers to support all of its clients.
- **Computers crash.** Since Reliable upgraded its operating system 12 months ago to JupiterOS v3.0, it has been experiencing a number of unexplained system crashes, averaging over 1 crash per week per computer, but with some variation over computers:

Computer:	A	B	C	D	E	F	G	H
Crashes/week:	1.2	2.0	0.6	0.8	1.4	1.2	1.0	0.4

This crash rate is barely acceptable.

- **New release, JupiterOS v3.1.** Supposed to greatly reduce the number of unexplained crashes. Reliable's systems managers are skeptical of such claims and worried that the costs involved in upgrading to v3.1 may not justify the increased reliability, or, worse, that the new version may be even more crash prone.

Experiment!

- **Upgrade only 2 of the 8 machines** so they can compare v3.0 and v3.1 over a number of weeks. Any other changes to the computers will be done identically to all 8.
- **How long should they test v3.1?**

Outcome Measure

- **Number of crashes per week.** Let Y_{it} be the number of crashes for Computer i in week t.
- **Y_{it} ~ Poisson(m_i).** Assume crashes are independent events. Then Y_{it} is distributed as a Poisson random variable with mean m_i.

Predictor variables

- **PastCrRt.** Past crash rates given above.
- **V31.** Code the dummy variable:
  V31 = 0, if running v3.0
  V31 = 1, if running v3.1

Assumed Model for Sample-Size Analysis

- Take the linear predictor to be

$$\ln\{m_i\} = 0.0 + 1.0*\ln[\text{PastCrRt}_i] - \ln(3)*V31_i.$$

  - Makes makes future crash rates for v3.0 equal to past crash rates.
  - But makes v3.1 have 1/3 the number of crashes as v3.0 (given the same past crash rate).

Steps to Perform Complex *yet Practical* Power Analyses

1. Formulate the problem completely. This is the hardest part.
2. Create exemplary data set for the problem. Straightforward programming.
3. "Analyze" exemplary data set using ordinary software. A dry run of the planned data analysis.
4. Convert test statistics to noncentrality values and then to powers (for given Ns) or sufficient Ns (for given powers). Easy with UnifyPow, a freeware SAS module/macro. (http://www.bio.ri.ccf.org/power.html)

2. Create exemplary data set for the problem.

- Via the SAS System.

```
data crashes;
title2 "Get exemplary LR test value for computer crashing problem";
keep PastCrRt V31 m Ncrashes lnPCR lnTotWks;
input Computer $ PastCrRt V31;
TotWeeks = 100;
betaInt = 0.0;
betaPCR = 1.0;
betaV31 = -log(3);  *Note: -ln(3) = = -1.0986;
m = exp(betaInt + betaPCR*log(PastCrRt) + betaV31*V31);
Ncrashes = TotWeeks*m;
lnPCR = log(PastCrRt);
lnTotWks = log(TotWeeks);


* Computer  PastCrRt  V31;      datalines;
     A        1.2      0
     B        2.0      1
     C        0.6      0
     D        0.8      1
     E        1.4      0
     F        1.2      0
     G        1.0      0
     H        0.4      0
```

Do not bother with during presentation!

- **The data set created.** 8 prototypical computers comprise the exemplary data set. The variable Ncrashes gives the exemplary number of crashes for 100 weeks.

Case	PastCrRt	v31	m	Ncrashes
1	1.2	0	1.20000	120.000
2	2.0	1	0.66667	66.667
3	0.6	0	0.60000	60.000
4	0.8	1	0.26667	26.667
5	1.4	0	1.40000	140.000
6	1.2	0	1.20000	120.000
7	1.0	0	1.00000	100.000
8	0.4	0	0.40000	40.000

3. "Analyze" exemplary data set using ordinary software.

- Key parts of input (SAS):

```
/*
   From data step:
                    TotWeeks = 100;
                    lnTotWks = log(TotWeeks);
                    lnPCR = log(PastCrRt);
*/
     proc genmod;
       model Ncrashes = lnPCR V31 /
          dist = Poisson
          link = log
          offset = lnTotWks
          type3;
```


- Key lines from output:

```
   LR Statistics For Type 3 Analysis

Source        DF   ChiSquare  Pr>Chi

   LNPCR       1    104.1605  0.0001
→  V31         1    104.5144  0.0001
```


4. Convert test statistics to powers or sufficient Ns.

```
title2 "Sample-size analysis for JupiterOS
v3.1 evaluation";
title3 "Units = weeks (all eight machines)";
%include "&UnifyPow";
datalines;
Exemplary chi**2
Nexemplary 100 .  weeks (all eight machines)
alpha .05 .10
power .90 .95
tails 2
effects
"v3.1 vs. v3.0"  1  104.5144
%tables;
```

Scenario: Exemplary chi**2

	ALPHA			
	0.05		0.1	
	Minimum Power		Minimum Power	
	.900	.950	.900	.950
	Total N	Total N	Total N	Total N
Effect	Statistic				
v3.1 vs. v3.0	2-tail Z	11	13	9	11

Just a Little Theory
(The proceedings paper will have a fuller discussion.)

Reference: Self, Mauritsen, and Ohara (Biometrics, 1992).

Consider any generalized linear model where the hypothesis of interest is formed by comparing a model with rank r_{full} parameters versus a nested model with $r_{reduced}$ parameters. The usual LR test statistic has a noncentral χ^2 distribution with $(r_{full} - r_{reduced})$ degrees of freedom and noncentrality parameter,

$$\lambda \cong N\lambda^* + \zeta$$

where λ^* and ζ are functions of the design matrix and true population parameters, **but not N.**

λ^* is relatively easy to understand and compute. (This what we have been doing.)

ζ is unwieldy, but fortunately it is usually inconsequential in value compared to $N\lambda^*$. ζ can be either positive or negative.

Thus in practice we can use:

$$\lambda \cong N\lambda^*$$

How much difference does it make to leave out ζ?

Logistic Regression Example ($\alpha = .01$)

Total N	Nominal Power (2-sided test)		95% Confidence Limits (5000 simulations)		$\frac{\zeta}{N\lambda^* + \zeta}$
	$N\lambda^* + \zeta$	$N\lambda^*$	lower	upper	
222	.806	.802	.791	.813	.0083
282	.903	.901	.898	.915	.0066
339	.952	.951	.954	.964	.0055

Poisson Regression Example ($\alpha = .10$)

Total Weeks	Nominal Power (2-sided test)		95% Confidence Limits (5000 simulations)		$\frac{\zeta}{N\lambda^* + \zeta}$
	$N\lambda^* + \zeta$	$N\lambda^*$	lower	upper	
9	.928	.923	.933	.946	.0251
11	.962	.960	.969	.978	.0205

Thanks for coming!

Go Tribe!

Power for Categorical Data Analysis: What Problems Do Some General Tools Handle?

Problem	nQuery Advisor 2.0	Power & Precision a.k.a SamplePower	PASS 6.0	UnifyPow.sas
compare 2 groups on ordered categorical variable using W-M-W rank-sum	Noether power approximation (less accurate?); balanced design only			Lehmann-Hettesmansperger (default), Noether, and ARE-based power approximations; also gives "lower bound" on power via ARE argument
kappa for 2 x 2	yes			
beta-binomial problem: comparing G groups on a beta "latent" variable, π_{ij}, having means $\mu_j(\pi_{ij})$ and standard deviation $\sigma(\pi_{ij})$ data are binomial(N, π_{ij}).				added for particular CCF project; untested approximation; in code but not yet documented
logistic regression	Whittemore/Hsieh method for single beta (narrow range of applicability?)		Whittemore/Hsieh method for single beta (narrow range of applicability?)	great generality, but two-step process
generalized linear models				great generality, but two-step process

Power for Categorical Data Analysis: What Problems Do Some General Tools Handle?

Problem	nQuery Advisor 2.0	Power & Precision a.k.a SamplePower	PASS 6.0	UnifyPow.sas
1 proportion versus null	standard Z approximation	Arcsin approx (default); exact available	exact	exact
Comparing 2 independent proportions, i.e., assoc in 2 x 2 table	Pearson χ^2; Fisher's exact conditional	Pearson χ^2; arcsin method; Fisher's exact cond'l	Pearson χ^2; arcsin method; Fisher's exact cond'l	Pearson χ^2; exact *unconditional;* Fisher's exact cond'l; likelihood ratio (LR)
McNemar's test comparing 2 correlated proportions; i.e. the off-diagonal cells of the 2 x 2 table	Miettinen's approx calculation for power	"naive" approx calculation for power (may work OK)	Instead: "Matched Case-Control" problem	exact power calculation
Compare >2 proportions, i.e., assoc in G x 2 table	Overall Pearson χ^2; linear trend on *logits*	Overall Pearson χ^2		Overall Pearson χ^2; Overall LR; general Wald-type contrasts on *logits*, including $df_H > 1$
Overall association in G x C contingency table		Pearson χ^2		Pearson χ^2; likelihood ratio

Ralph G. O'Brien, Cleveland Clinic Foundation
1998 Joint Statistical Meetings (Dallas)

For information on the SAS-based freeware, UnifyPow, visit
http://www.bio.ri.ccf.org/power.html