

Sample size calculations

SAS proc power
(sort of parallel to HL Chapter 10.5)

The “Rule of 10”

- The number of model covariates should not exceed (by much) the frequency of the least frequent outcome divided by 10
- Example:
 - ICU study: $n=200$, $STA=1$ for 40 observations, $STA=0$ for 160 observations
 - # of covariates $\leq 40/10 \approx 4$

Sample size calculations

- Example:
 - Assume the Glow500 data set is from a pilot study
 - Assume we want to calculate the required sample size for a larger study
- Recall: Outcome = FRACTURE

Risk factors of interest

- Risk factors of interest
 - MOMFRAC
 - AGE
- Estimate sample size to detect the effect of
 - MOMFRAC on FRACTURE, if it exists
 - AGE on FRACTURE, if it exists

Potential confounders

- Want to estimate power univariately and multivariately
- Suspected confounders are
 - PRIORFRAC
 - ARMASSIST
 - RATERISK2
 - AGE
 - HEIGHT

MOMFRAC

- Univariate analysis

Sample size in OpenEpi

Sample Size for Unmatched Case-Control Study

Two-sided confidence level(1-alpha)	95
Power(% chance of detecting)	80
Ratio of Controls to Cases	3
Hypothetical proportion of controls with exposure	10
Hypothetical proportion of cases with exposure:	18.18
Least extreme Odds Ratio to be detected:	2.00

Where did the entries come from?

- Usually choose $\alpha = 0.05$
- Usually chose power=80%
- Pilot study: 125 cases, 375 controls (3 times as many controls as cases)
- Pilot study: About 10% of controls had MOMFRAC=yes

MOMFRAC	FRACTURE=yes		FRACTURE=no	
	Frequency	Percent	Frequency	Percent
Yes	24	19.2	41	10.9
No	101	80.8	334	89.1
Total	125	100.0	375	100.0

- Assume we want to detect an OR of 2.0

OpenEpi Results

	Kelsey	Fleiss	Fleiss with CC
Sample Size - Cases	166	179	195
Sample Size - Controls	497	535	583
Total sample size:	663	714	778

- We need a total sample size of about 800 with about 200 cases (25%) and 600 controls (75%)

Sample size in SAS

```
proc power;
  logistic
    vardist("momfrac") = binomial(.1,1)
    testpredictor = "momfrac"
    responseprob = 0.25
    testoddsratio = 2.0
    ntotal = .
    power = 0.8;
run;
```

What does all this mean?

`vardist("momfrac") = binomial(.1,1)` 10% of controls have MOMFRAC=yes

`testpredictor = "momfrac"` Variable of interest is MOMFRAC

`responseprob = 0.25` 25% of the sample are cases

`testoddsratio = 2.0` We want to detect an OR of 2.0

What does all this mean?

`ntotal = .` We want to estimate the sample size

`power = 0.8;` We want 80% power

$\alpha = 0.05$ is the default and does not need to be specified

SAS Results

Fixed Scenario Elements

Method	Shieh-O'Brien approximation
Response Probability	0.25
Test Predictor	momfrac
Odds Ratio for Test Predictor	1.5
Unit for Test Pred Odds Ratio	1
Nominal Power	0.8
Total Number of Bins	2
Alpha	0.05

SAS Results

Computed N Total

Actual Power	N Total
0.800	832

- We need a total sample size of about 800 with about 200 cases (25%) and 600 controls (75%)

MOMFRAC

- Multivariate analysis

Sample size from SAS

```
proc power;
  logistic
```

```
    vardist ("momfrac") = binomial(0.1, 1)
    vardist ("priorfrac") = binomial(0.2, 1)
    vardist ("armassist") = binomial(0.3, 1)
    vardist ("raterisk2") = binomial(0.25, 1)
```

```
    vardist ("age") = normal(67.5, 9)
    vardist ("height") = normal(162, 6)
```

Sample size from SAS

```
testpredictor = "momfrac"
covariates = ("priorfrac" "armassist" "raterisk2"
             "age" "height")
responseprob = 0.25
testoddsratio = 2.0
covoddsratios = (2.0 1.5 1.5 1.5 0.6)
units = ("age"=10 "height"=10)
ntotal = .
power = 0.8;
```

```
run;
```

What does all this mean?

```
vardist ("momfrac") = binomial(0.1, 1)
vardist ("priorfrac") = binomial(0.2, 1)
vardist ("armassist") = binomial(0.3, 1)
vardist ("raterisk2") = binomial(0.25, 1)
```

From Chapter4_4Results.pdf, Table 1:

```
≈10% of controls have momfrac=yes
≈20% of controls have priorfrac=yes
≈30% of controls have armassist=yes
≈25% of controls have raterisk2=yes
```

What does all this mean?

```
vardist ("age") = normal(67.5,9)
vardist ("height") = normal(162,6)
```

From Chapter4_4Results.pdf, Table 2:

Age: Among controls, mean=67.5, std ≈9
Height: Among controls, mean=162, std ≈6

Continuous variables may have a distribution other than normal (e.g. lognormal)

What does all this mean?

testpredictor = "momfrac" Variable of interest is MOMFRAC

covariates = ("priorfrac" "armassist" "raterisk2" "age" "height") Additional variables in the model

responseprob = 0.25 25% of the sample are cases

testoddsratio = 2.0 We want to detect an OR of 2.0

What does all this mean?

covoddsratios = (2.0 1.5 1.5 1.5 0.6)

- Estimated ORs for the other model covariates (in the order listed above)
- Estimates are from Chapter4_4Results.pdf, Table 7

units = ("age"=10 "height"=10)

- The ORs for age and height reflect unit increases of 10

What does all this mean?

ntotal = . We want to estimate the sample size

power = 0.8; We want 80% power

$\alpha = 0.05$ is the default and does not need to be specified

Results from SAS

Fixed Scenario Elements	
Method	Shieh-O'Brien approximation
Response Probability	0.25
Test Predictor	momfrac
Odds Ratio for Test Predictor	2
Unit for Test Pred Odds Ratio	1
Covariates	priorfrac armassist raterisk2 age height
Covariate Odds Ratios	2 1.5 1.5 1.5 0.6
Units for Covariate Odds Ratios	1 1 1 10 10
Nominal Power	0.8
Total Number of Bins	1600
Alpha	0.05

What do the results tell us?

Computed N Total	
Actual Power	N Total
0.800	879

- We need a total sample size of about 900 with about 225 cases (25%) and 675 controls (75%)

AGE

- Univariate analysis

Sample size from SAS

```
proc power;  
  logistic  
    vardist("age") = normal(67.5,9)  
    testpredictor = "age"  
    responseprob = 0.25  
    testoddsratio = 1.5  
    units = ("age"=10)  
    ntotal = .  
    power = 0.8;  
run;
```

We want to detect an OR of 1.5 for a 10 year increase in age

Results from SAS

Fixed Scenario Elements	
Method	Shieh-O'Brien approximation
Response Probability	0.25
Test Predictor	age
Odds Ratio for Test Predictor	1.5
Unit for Test Pred Odds Ratio	10
Nominal Power	0.8
Total Number of Bins	10
Alpha	0.05

What do the results tell us?

Computed N Total	
Actual Power	N Total
0.801	361

- We need a total sample size of about 350 with about 90 cases (25%) and 260 controls (75%)

AGE

- Multivariate analysis

Sample size from SAS

```
proc power;  
  logistic  
    vardist ("momfrac") = binomial(0.1, 1)  
    vardist ("priorfrac") = binomial(0.2, 1)  
    vardist ("armassist") = binomial(.3, 1)  
    vardist ("raterisk2") = binomial(0.25, 1)  
    vardist ("age") = normal(67.5, 9)  
    vardist ("height") = normal(162.6)
```

Sample size from SAS

```
testpredictor = "age"
covariates = ("priorfrac" "armassist" "raterisk2"
             "momfrac" "height")
```

```
responseprob = 0.25
testoddsratio = 1.5
covoddsratios = (2.0 1.5 1.5 2.0 0.6)
units = ("age"=10 "height"=10)
ntotal = .
power = 0.8;
```

```
run;
```

Results from SAS

Fixed Scenario Elements	
Method	Shieh-O'Brien approximation
Response Probability	0.25
Test Predictor	age
Odds Ratio for Test Predictor	1.5
Unit for Test Pred Odds Ratio	10
Covariates	priorfrac armassist raterisk2 momfrac height
Covariate Odds Ratios	2 1.5 1.5 2 0.6
Units for Covariate Odds Ratios	1 1 1 1 10
Nominal Power	0.8
Total Number of Bins	1600
Alpha	0.05

What do the results tell us?

Computed N Total	
Actual Power	N Total
0.801	371

- We need a total sample size of about 375 with about 95 cases (25%) and 280 controls (75%)

Conclusion

- Would probably choose a sample size of about 900 (225 cases and 675 controls)
- It is the greatest sample size calculated
- It is needed to detect an OR of 2.0 for MOMFRAC in a multivariate model also containing PRIORFRAC, ARMASSIST, RATERISK2, AGE AND HEIGHT)

Problem

- The data needed to conduct the sample size calculations are not available unless a pilot study was conducted
- Maybe able to get estimates from the literature
- May conduct sensitivity analyses (i.e. obtain results for different sets of input values)