Power and Sample Size Calculations

Tyler J. VanderWeele
Departments of Epidemiology and Biostatistics
Harvard T.H. Chan School of Public Health

Plan of Presentation

(1) Sample Size and Power Calculations

(2) Observational Studies vs. RCTs

(3) Some Software Tools and Additional Resources

(4) Sample Size and Power for Cluster RCTs

(5) Sample Size and Power for Interactions

(6) Sample Size for Mediation

(7) Sample Size for Confidence Intervals

(8) Concluding Remarks

Sample Size and Power

Sample size calculations are an essential part of planning

Sample size requirements should be considered early in the planning phase of a study

A well-designed trial will be large enough to detect clinically important differences between groups with high probability

To perform sample size calculations, we need well defined study endpoints, hypotheses, and statistical tests

We also need estimates of the variability of the outcomes

Both analytically and empirically coming up with estimates of the variance is often the most challenging step

Sample Size and Power

Suppose we have a continuous outcome Y and wish to test the null hypothesis:

H_0: E[Y|A=1] = E[Y|A=0] i.e. no treatment effect

Versus the alternative hypothesis:

H_A: E[Y|A=1] ≠ E[Y|A=0] i.e. non-zero treatment effect

In a randomized trial with equally sized groups we can do so by specifying:

(i) The significance level α (e.g. 5%) and power P=1-β (e.g. P=80%)
(ii) The magnitude of the effect η = E[Y|A=1] - E[Y|A=0] that we wish to be able to detect

Sample Size and Power

To calculate the sample size for *each* group required to detect a treatment effect size η using a t-test with significance level α and power *P* (=1-Type II error (β), i.e. *P*=1-β) we can use:

$$N=(Z_{1-a/2} + Z_P)^2 \times V \ / \ \eta^2$$

where $Z_{1-a/2}$ and Z_P denote respectively the (1-α/2)th and *P*th quantiles of a standard normal variable (e.g. $Z_{1-a/2}$=1.96 for α=5% and Z_P=0.84 for *P*=80%)

And where if the variance of Y is assumed to be the same in the treatment and the control groups then $V=2\sigma^2$ and σ^2 is the variance of Y

Other formulas are available if the variance is different across groups or if treated and control groups are different sizes

Sample Size and Power

Example: Suppose we were testing a drug to lower LDL cholesterol and wanted to know the sample size required to detect an effect size of 13 point decline with power=80% and we thought the standard deviation of the outcome for the study population was 35 then we could use for each group:

$$N = (Z_{1-\alpha/2} + Z_P)^2 \times V \; / \; \eta^2$$

$$N = (1.96 + 0.84)^2 \times (2*35^2) \; / \; (13)^2$$

$$N = 114$$

Sample Size and Power

The difficult part of obtaining the information necessary to undertake sample size calculations is generally getting an estimate of the variance

This can be done in a few different ways:

1. Search the literature
2. Use approximate methods to estimate the variance
   e.g. for approximately normal variables, the standard deviation is roughly ¼ of the range of the 95% most common values
3. Conduct a pilot study
4. Dichotomize the variable (see below)
5. Make an educated guess
6. Consider multiple plausible scenarios

Sample Size and Power

If instead of calculating sample size, we had a fixed sample size N in each group and wanted to calculate power to detect a treatment effect size of magnitude η the following formula can be used:

$$\text{Power} = \Phi\{ - Z_{1- a/2} + \eta \sqrt{N / V} \}$$

where:

$Z_{1- a/2}$ is the $(1-\alpha/2)$th quantile of a standard normal variable (e.g. $Z_{1- a/2}=1.96$ for $\alpha=5\%$)

Φ is the cumulative distribution function for a standard normal random variable

$V=2\sigma^2$ and σ^2 is the variance of Y

Sample Size and Power

For a **binary** outcome Y, to detect a treatment effect size of η=P(Y=1|A=1) - P(Y=1|A=0), we can use the same formulae for sample size and power calculations but simply use a different value of V

For sample size in group, we thus have:

$$N=(Z_{1-\,a/2} + Z_P)^2 \times V \;/\; \eta^2$$

For power, if we have N in each group then the power is:

$$\text{Power} = \Phi\{ - Z_{1-\,a/2} + \eta \sqrt{(N / V)} \}$$

where: $V = p_0(1-p_0) + p_1(1-p_1)$ and p_0=P(Y=1|A=0), p_1=P(Y=1|A=1) under the alternative

or where: V = 2p(1-p) and where p is the overall pooled probability of an event i.e. p=P(Y=1) under the null

Sample Size and Power

In some studies, it may be that due to limited resources or limited number of available patients, the sample size is infeasible

In such settings there are several things to consider

1. Is the effect size unreasonably small or the assumed variance too large?
2. Is it reasonable to increase Type 1 error or decrease power?
3. Choose continuous primary outcomes
4. Choose outcomes that have smaller variance
5. Consider surrogate variables
6. Choose binary outcomes that are more common

However, all of these approaches are potentially problematic and it is important to be realistic

An underpowered study is a bad investment for everyone

Sometimes it is best not to undertake the study or to wait until more resources or patients are available

Sample Size and Covariates

In a randomized trial control for covariates predictive of the outcome will increase power

One's sample size and power calculations will thus be conservative
Perhaps still best to use power calculations without the covariate control because of

- drop-out
- tendency to be over-optimistic about effect size / variance
- data may be used for other purposes (subgroup analysis)

Tests using linear and logistic regression in RCTs are robust to model mis-specification (Rosenblum and van der Laan, 2009)
Improving efficiency of estimates using baseline covariates while maintaining robustness to misspecfication is an active area of research (Tsiatis et al., 2008; Zhang et al., 2008; Lu et al., 2008)

Observational Studies and Covariates

In observational studies, covariate control is needed for confounding
Effects on power / sample size can go in either direction

If the covariates are predictive of the outcome, power can improve
If more predictive of the exposure, power can decline
Confounding can alter the effect size in either direction

It is often difficult to establish this in advance
Simulations can sometimes help assess how power is affected

With control for numerous correlated covariates, often power goes down
In planning one may want to be even more conservative in observational studies
But often observational studies use secondary data and sample size is fixed

Power for Given Sample Size

	Standardized Effect Size (σ^2=1 and V=2σ^2=2)			
	η=0.1	η=0.15	η=0.2	η=0.3
Total N=500 (250/group)	0.2	0.30	0.61	0.91
Total N=1000 (500/group)	**0.35**	**0.66**	**0.89**	**0.997**
Total N=2000 (1000/group)	0.61	0.91	0.994	0.999

Moreover, these calculations optimistically assume:

- Equal sample sizes for each group
- No power loss due to covariate control in observational studies
- No missing data (attenuating sample size)
- No measurement error (attenuating effect estimates)

Power even with Total N=1000 is only moderate for small effect sizes
Should we be restricting attention to studies with N>1000…?

Sample Size and Power

Some relevant further literature (see next slide for references) for power and sample size calculations include:

Outcome	Test	Reference	Resource Type
Binary	χ^2	1	Tables
	Normal	2	Formulae
	Fisher's	3	Tables/F
Continuous	Nonparametric	4	Tables
	F-test	5	Formulae
		6	Tables
		7	Program
		8	Graphs
Survival Time	Logrank	9	Formulae
		10	Tables

Sample Size and Power

1. Fleiss JL. *Statistical Methods for Rates and Proportions.* New York: Wiley, 1981. pp 38-48.

2. Feigl P. A graphical aid for determining sample size when comparing two independent proportions. *Biometrics.* 1978; 34:111-122.

3. Casagrande JT, Pike MC, Smith PG. The power function of the exact test for comparing two binomial distributions. *Appl Stat.* 1976; 27;176-180.

4. Lehmann EL. *Nonparametrics: Statistical Methods Based on Ranks.* San Francisco: Holden-Day, 1975.

5. George SL, Desu MM. Planning the size and duration of a clinical trial studying time to some critical event. *J Chron Dis.* 1974; 27:15-24.

6. Lesser ML, Cento SJ. Tables of power for the F-test for comparing two exponential survival distributions. *J Chron Dis.* 1981; 34-533-544.

7. Bernstein D, Lagakos SW. Sample Size and power determination for stratified clinical trials. *J Stat Comput Simul.* 1978; 8:65-73.

8. Schoenfield DA, Richter JR. Monograms for calculating the number of patients needed for a clinical trial with survival as an endpoint. *Biometrics.* 1982; 38:163-170.

9. Peto R, Pike MC, Armitage P, et al. Design and analysis of clinical trials requiring prolonged observation of each patient. *Br J Cancer. Part I* – 1976; 34:585-612. Part II – *1977; 35:1-39.*

10. Freedman LS. Tables of the number of patients required in clinical trials using the logrank test. *Stat in Med.* 1982; 1:121-129.

Sample Size and Power

Some relevant web resources include:

UCLA http://calculators.stat.ucla.edu/powercalc/

- Normal, Exponential, Binomial, Poisson
- Sample Size and Power Calculations

MGH: http://hedwig.mgh.harvard.edu/sample_size/size.html

- Binomial (Parallel or Cross-over Trial)
- Normal (Parallel or Cross-over Trial)
- Time to Event

Statpages: http://statpages.org/

- An excellent compendium of online statistical tools

Cluster Randomized Trial

Cluster Randomized Trials

- Individuals are clustered in group
- The entire group is randomized to treatment or control
- Preserves the advantages of randomization
- But requires special care in design and analysis

Reasons for a Cluster Randomized Trial:

- Administrative convenience
- To avoid treatment group contamination
- Intervention is naturally applied at the cluster level

Cluster Randomized Trial

Example:

- Vaccine Trial
- There could be interference/contamination in an individually randomized trial

Example:

- Conflict Resolution RCT in Schools (e.g. Jones et al, 2010)
- There would be interference in an individually randomized trial
- Also it would be less practical to implement

Implications of Clustering

- A key property of cluster randomization trials is that inferences are frequently intended to apply at the individual level while randomization is at the cluster or group level. Thus the unit of randomization may be different from the unit of analysis.

- In this case, the lack of independence among individuals in the same cluster, i.e. between-cluster variation, creates special methological challenges in both design and analysis.

Implications of Clustering

(i) Reduction in effective sample size.

- Extent depends on degree of within-cluster correlation and on average cluster size

(ii) Standard approaches for sample size estimation and statistical analysis do not apply.

- Application of standard sample size approaches leads to an underpowered study.
- Random effects or generalized estimating equations are often used to deal with clustering in analysis
- Application of standard statistical methods generally tends to underestimate standard errors

Effective Sample Size

Consider a trial in which k clusters of size m are randomly assigned to each of an experimental and control group. Also assume the response variable Y is normally distributed with common variance σ^2

Aim is to test $H_0: \mu_1 = \mu_2$ where μ_1 and μ_2 are the means with and without treatment.

Because of clustering the variance will be inflated

The term IF = 1 + (m- 1) ρ is the "variance inflation factor" or "design effect" associated with cluster randomization

where ρ is the coefficient of intracluster correlation.

Effective Sample Size

The overall response variance σ^2 may be expressed as the sum of two components, i.e.,

$$\sigma^2 = \sigma^2_A + \sigma^2_W ,$$

where

σ^2_A = between-cluster component of variance

σ^2_W = within-cluster component of variance

then

$$\rho = \sigma^2_A / (\sigma^2_A + \sigma^2_W)$$

Effective Sample Size

Comparison of Means:

Suppose *k* clusters of size *m* are to be assigned to each of two intervention groups.

Then the number of subjects required per intervention group at significance α and power *P* to test $H_0: \mu_1 = \mu_2$ is given by

$$N = (Z_{1-a/2} + Z_P)^2 \times V \; / \; \eta^2$$

Where $V = (2\sigma^2)\,[1 + (m - 1)\,\rho]$

And $\sigma^2 = \sigma^2_A + \sigma^2_W$ $\qquad\qquad$ $\eta = (\mu_1 - \mu_2)$

The number of required clusters per intervention group is given by $k = N/m$.

Example

Hsieh (1988) reported on the results of a pilot study for a planned trial of work-site intervention to improve cardiovascular risk factors; one outcome of interest was cholesterol levels; pilot data was obtained from 754 individuals in 4 worksites.

Estimated variance components were

$$S^2_W = 2209,\ S^2_A = 93,\ \sigma^2 = 2302\ (\sigma = 47.98)$$

$\therefore$ value of ρ assessed as $\rho = 93 / (93 + 2209) = 0.04$

Assuming = 70 subjects/worksite,

$$IF = 1 + (70 - 1)\ 0.04 = 3.76$$

Example

To obtain 80% power at α = .05 (2 sided) for detecting a mean difference of 20 mg/dl between intervention groups, the number of required worksites per treatment group is given by

$$n = \{(Z_{1-\alpha/2} + Z_P)^2 (2\sigma^2) [1 + (m - 1) \rho]\} / (\mu_1 - \mu_2)^2$$

$$n = \{(1.96 + 0.84)^2 2(2302) (3.76)\} / (20)^2 = 339$$

$$k = n/m = 339/70 = 4.8 \cong 5$$

Might want to increase this number for possible loss to follow-up

Effective Sample Size

Number of communities	Subjects/community	Total subjects	Design effect D (assuming ρ=0.017)	Effective sample size	Power (h=0.2 effect size)
4	320	1280	6.42	199	.514
8	160	1280	3.70	346	.749
16	80	1280	2.34	547	.911
32	40	1280	1.66	771	.975
64	20	1280	1.32	970	.993
128	10	1280	1.15	1113	.997
1280	1	1280	1	1280	.999

Practical Considerations

Trial randomizing between 30 and 50 individuals per cluster will tend to have almost the same statistical power as trials randomizing the same number of much larger units

- But clusters of larger size are often recruited for practical reasons

Another approach to cluster randomized trials is to pair match based on available covariate and then randomize

- This has the advantage of generally making treated and control groups more comparable
- When sample sizes of clusters dramatically differs this can be especially helpful and sample size of the cluster can be taken as a matching variable (Imai et al., 2009)

Sample Size for Interactions in Logistic Models

Consider the logistic regression model:

$\text{logit}(\text{pr}(D=1|G=g,E=e))=\beta_0+\beta_1 g+\beta_2 e+\beta_3 ge$

Suppose we were interested in the sample size required to detect a log interaction ratio of $\beta_3=\eta$

We can do so by specifying:

(1) The significance level α (e.g. 5%) and power *P* (e.g. 80%)

(2) The main effect log odds ratios β_1 and β_2

(3) The proportions in each GxE exposure strata $\pi_{ij}=P(G=i,E=j)$

Instead of the proportions π_{ij} we can alternative specify (i) the overall prevalence of G and E i.e $p_g=P(G=1)$, $p_e=P(E=1)$, (ii) the odds ratio relating G and E, $\Delta=\{P(G=1|E=1)/P(G=0|E=1)\}/\{P(G=1|E=0)/P(G=0|E=0)\}$ and (iii) $P(Y=1|G=0,E=0)$ [see Appendix slides]

This latter specification will be easier to use in a case control study

Note that $\beta_0=\text{logit}\{P(Y=1|G=0,E=0)\}$ where $P(Y=1|G=0,E=0)$ is the prevalence of cases in the doubly unexposed group in the study (e.g. in the actual case-control sample)

Sample Size for Interactions in Logistic Models

To calculate the overall sample size (cases and controls) required to detect a log interaction ratio of $\beta_3=\eta$ with significance level α and power *P* we can use:

$$N=(Z_{1-\,a/2} + Z_P)^2 \times V \;/\; \eta^2$$

where $Z_{1-\,a/2}$ and Z_P denote respectively the $(1-\alpha/2)$th and *P*th quantiles of a standard normal variable (e.g. $Z_{1-\,a/2}=1.96$ for $\alpha=5\%$ and $Z_P=0.84$ for $P=80\%$)

And where V is the variance of the estimator of β_3 evaluated under the alternative

This variance V takes the form (Demidenko, 2008):

$$V = \frac{1}{L} + \frac{1}{R} + \frac{1}{F} + \frac{1}{J}$$

Sample Size for Interactions in Logistic Models

This variance V takes the form: $$V = \frac{1}{L} + \frac{1}{R} + \frac{1}{F} + \frac{1}{J}$$

Where:

$$L = \frac{\exp(\beta_0)}{[1+\exp(\beta_0)]^2}\ \pi_{00}$$

$$F = \frac{\exp(\beta_0+\beta_1)}{[1+\exp(\beta_0+\beta_1)]^2}\ \pi_{10}$$

$$J = \frac{\exp(\beta_0+\beta_2)}{[1+\exp(\beta_0+\beta_2)]^2}\ \pi_{01}$$

$$R = \frac{\exp(\beta_0+\beta_1+\beta_2+\eta)}{[1+\exp(\beta_0+\beta_1+\beta_2+\eta)]^2}\ \pi_{11}$$

Sample Size for Interactions in Logistic Models

If instead we want to calculate power to detect a log interaction ratio of magnitude η from a given dataset with sample size N the following formula can be used:

$$\text{Power} = \Phi\{ - Z_{1- a/2} + \eta \sqrt{N / V} \}$$

Where:

$Z_{1- a/2}$ is the $(1-\alpha/2)$th quantile of a standard normal variable (e.g. $Z_{1- a/2}=1.96$ for $\alpha=5\%$)

Φ is the cumulative distribution function for a standard normal random variable

V is obtained by the expressions given above

Sample Size for Interactions in Logistic Models

These can all be done by hand

Alternatively there is software available for this at:

www.dartmouth.edu/~eugened

Demidenko's spreadsheet is relatively straightforward for cohort studies

However, for case-control data it requires specifying P(Y=1|G=0,E=0) i.e. the prevalence of cases in the doubly unexposed group in the case-control sample

This is not a particularly straightforward quantity to specify

We could obtain it by specifying the proportion of cases to control, along with p_g=P(G=1), p_e=P(E=1), and the odds ratio relating G and E, but this then requires solving a 'quartic' equation (Demidenko, 2008)

Alternatively, under a rare outcome assumption, we can obtain simple analytic expression (VanderWeele, 2012)

Excel spreadsheet to implement this automatically

Sample Size for Interactions in Logistic Models

Prior to Demidenko (2008), there had been a number of previous sample size calculations for interactions in logistic regression (Hwang et al., 1994; Foppa and Spiegelman, 1997)

However, these calculated the variance under the null rather than the alternative (Garcia-Closas and Lubin, 1999; Demidenko, 2008) which led to underestimates of the sample size when the interaction ratio was relatively large if variance is calculated under the alternative (as it is in nearly all software!)

The sample size calculation above are for Wald-tests (standard SAS/Stata/R output) with variance under the alternative

Lubin and Gail (1990) describe a method for the score test with variance under the alternative

Gauderman et al. (2002ab) developed sample size calculations for various matched designs and family-based designs us likelihood ratio tests

Sample Size for Additive Interaction

Consider again the logistic regression model:

$$\text{logit}(\text{pr}(D=1|G=g,E=e))=\gamma_0+\gamma_1 g+\gamma_2 e+\gamma_3 ge$$

Suppose we were interested in the sample size required to detect additive interaction of magnitude RERI=η

We can again do so by specifying:
(1) The significance level α (e.g. 5%) and power *P* (e.g. 80%)
(2) The main effect log odds ratios γ_1 and γ_2
(3) The proportions in each GxE exposure strata π_{ij}=P(G=i,E=j)

Or, instead of (3) by specifying:
(i) the overall prevalence of G and E i.e p_g=P(G=1), p_e=P(E=1),
(ii) the odds ratio relating G and E
(iii) The proportion of cases to controls in the study

Sample Size for Additive Interaction

To calculate the overall sample size required to detect additive interaction of RERI=η with significance level α and power *P* we can use:

$N=(Z_{1-\,a/2} + Z_P)^2 \times V \;/\; \eta^2$

where $Z_{1-\,a/2}$ and Z_P denote respectively the (1-α/2)th and *P*th quantiles of a standard normal variable (e.g. $Z_{1-\,a/2}$=1.96 for α=5% and Z_P=0.84 for *P*=80%)

And where V is the variance of the estimator RERI given by (VanderWeele, 2012):

$$V_{RERI(OR)} = (\frac{1}{L}+\frac{1}{R})e^{2(\gamma_1+\gamma_2+\gamma_3)} - \frac{2}{L}e^{2\gamma_1+\gamma_2+\gamma_3} - \frac{2}{L}e^{\gamma_1+2\gamma_2+\gamma_3} + (\frac{1}{L}+\frac{1}{F})e^{2\gamma_1} + (\frac{1}{L}+\frac{1}{J})e^{2\gamma_2} + \frac{2}{L}e^{\gamma_1+\gamma_2}$$

with L, R, F, J as given above and in Demidenko (2008)

As before power for a given sample size N is obtained by:
Power = $\Phi\{ - Z_{1-\,a/2} + \eta \sqrt{N / V} \}$

Sample Size for Additive Interaction

Similar expressions (VanderWeele, 2012) are available for sample size or power if…

(1) RERI is calculated from a log-linear model rather than a logistic model
(2) RERI is calculated using logistic regression with case-control data assuming a rare outcome
(3) Additive interaction using risks are used

Slightly different expressions are needed to estimate π_{ij} which are needed to calculate L, R, F, J which are needed to obtain the variance V for RERI (VanderWeele, 2012)

Sample Size for Interaction

Excel spreadsheets are available (VanderWeele, 2012) to calculate power and sample size for additive interaction for either cohort studies (using risks or RERI) or for case control studies (using RERI) or for multiplicative interaction and for case-only estimators of interaction and these comparisons can be informative [see Appendix slides]

Power is often very low for interaction analyses
There will likely be many false positives (as has played out in GxE literature)
Rule of Thumb: 4 Times the Sample Size is required to detect interaction

This has implications for when one ought to undertake an interaction analysis
Often it will be inappropriate if a study is only powered to detect a main effect
One needs caution with regard to RCT subgroup analyses, especially if these are not specified a priori

Sample Size for Interaction

Table 1. Power to detect additive interaction and multiplicative interaction for various sample sizes, main effects, and interaction parameters (first number in each column is power to detect additive interaction; second number is power for multiplicative interaction)

I_{OR}	OR_{10}	OR_{01}	$n = 500$	$n = 1000$	$n = 3000$	$n = 5000$
1.1	1	1	.05, .05	.06, .06	.10, .09	.14, .13
1.1	1.3	1.3	.07, .04	.10, .05	.23, .09	.34, .12
1.1	1.5	1.8	.13, .04	.23, .05	.55, .08	.77, .11
1.3	1	1	.12, .11	.21, .17	.50, .42	.72, .62
1.3	1.3	1.3	.18, .10	.32, .15	.73, .37	.91, .56
1.3	1.5	1.8	.27, .09	.48, .14	.91, .33	.99, .50
1.5	1	1	.25, .19	.44, .34	.88, .77	.98, .93
1.5	1.3	1.3	.32, .17	.56, .30	.95, .70	1.00, .89
1.5	1.5	1.8	.40, .15	.68, .26	.99, .63	1.00, .84
2	1	1	.57, .44	.85, .73	1.00, .99	1.00, 1.00
2	1.3	1.3	.58, .39	.86, .65	1.00, .98	1.00, 1.00
2	1.5	1.8	.59, .34	.87, .59	1.00, .97	1.00, 1.00
3	1	1	.81, .77	.98, .97	1.00, 1.00	1.00, 1.00
3	1.2	1.3	.74, .70	.96, .94	1.00, 1.00	1.00, 1.00
3	1.5	1.8	.68, .62	.93, .89	1.00, 1.00	1.00, 1.00

In general whenever main effects and the interaction is positive, there will be more power to detect additive interaction than multiplicative interaction…

especially when the interaction is modest but the main effects are large

When outcome probabilities are additive or sub-additive power for multiplicative interaction is sometimes greater (Greenland, 1983)

Sample Size for Mediation

Very little formal work on sample size calculations for direct and indirect effects

Many papers use and cite simulation results from:
Fritz MS, MacKinnon DP. (2007) Required Sample Size to Detect the Mediated Effect. Psychological Science 18:193-198.

Helpful Rule of Thumb (Kenny and Judd, 2014):
In many (but not all!) contexts, power to detect indirect effect is greater than the total effect; but power to detect the direct effect is less than the total effect
Can be helpful in planning mediation analyses
Also, because of power, one must be careful of claims of “no direct effect” and therefore “complete mediation”

There has been some recent work and tools on simulation-based approaches for mediation power and sample size calculations (Rudolph et al., AJE 2020)

Hypothesis Testing vs. CI's

Power and sample size calculations often presume a hypothesis testing (p=0.05) framework; this can be problematic as we will see in the replication lecture

Should we do sample size for the confidence interval length instead...? (Rothman and Greenland, 2018)

For a Continuous Outcome (with N in *each* of two groups):

95% Confidence Interval is: Estimate +/- 1.96 x sqrt($2\sigma^2/N$)

Total Length of Interval is: 2 x 1.96 x sqrt($2\sigma^2/N$)

Solving for N gives: $N = 30.73\ \sigma / \text{Length}^2$

For $\sigma^2=1$ to get CI to be (Estimate +/- 0.1) we would need:

$N = 30.73\ \sigma / (0.2)^2 = 768$ in *each* group (i.e. 1536 total)

If $\sigma^2=1$, and we have N=500 in each group (1000 total) then we have:

95% CI = (Estimate - 0.12, Estimate + 0.12)

See Rothman and Greenland (2018) for formulas for binary outcomes

Conclusions

(1) Sample size and power calculations are essential in study planning

(2) Calculations somewhat less relevant with secondary data analyses when sample size is already fixed but still useful to see whether one should even bother and the extent we should worry about false positives

(3) Correlation needs to be taken into account in cluster RCTs

(4) Power for interaction analyses is often very low

(5) Power for direct effects is typtically lower than for indirect effects

(6) Power and sample size calculations often presume a hypothesis testing (p=0.05) framework; this can be problematic as we will see in the replication lecture; should we do sample size for the C.I. instead…?

Additional Slides

Tyler J. VanderWeele
Departments of Epidemiology and Biostatistics
Harvard T.H. Chan School of Public Health

Sample Size for Interactions in Logistic Models

If instead of the proportions π_{ij}=P(G=i,E=j) we specify (i) the overall prevalence of G and E i.e p_g=P(G=1), p_e=P(E=1), (ii) the odds ratio relating G and E, Δ={P(G=1|E=1)/P(G=0|E=1)}/{P(G=1|E=0)/P(G=0|E=0} and (iii) P(Y=1|G=0,E=0)

We can calculate π_{ij} as follows:

$\pi_{00} = (1 - p_e)/(1+C)$

$\pi_{01} = p_e/(1+C\Delta)$

$\pi_{10} = C(1 - p_e)/(1+C)$

$\pi_{11} = p_eC\Delta/(1+C\Delta)$

where $C= [q + sqrt\{q^2+4p_g(1- p_g)\Delta\}] / [2(1- p_g)\Delta]$

$q = p_g(1+\Delta) + p_e(1-\Delta) - 1$

Sample Size for Case-Only Interactions

Now consider a log-linear model
$\log(pr(D=1|G=g,E=e))=\beta_0+\beta_1 g+\beta_2 e+\beta_3 ge$

And the case-only estimator of β_3 under GxE independence:
$\beta_3= P(G=1|E=1,D=1)/P(G=0|E=1,D=1)\}/\{P(G=1|E=0,D=1)/P(G=0|E=0,D=1\}$

Suppose we were interested in the sample size required to detect a log interaction ratio of $\beta_3=\eta$

We can do so by specifying:
(1) The significance level α (e.g. 5%) and power *P* (e.g. 80%)
(2) The main effect risk ratios for the genetic factor alone $R_g=\exp(\beta_1)$ and for the environmental factor alone $R_e=\exp(\beta_2)$
(3) The overall proportions of G and E in the population; let g=P(G=1), e=P(E=1)

Sample Size for Case-Only Interactions

To calculate the sample size required to detect a log interaction ratio of $\beta_3=\eta$ with significance level α and power *P* we can once again use:

$$N=(Z_{1-\,a/2} + Z_P)^2 \times V \;/\; \eta^2$$

where $Z_{1-\,a/2}$ and Z_P denote respectively the $(1-\alpha/2)$th and *P*th quantiles of a standard normal variable (e.g. $Z_{1-\,a/2}=1.96$ for $\alpha=5\%$ and $Z_P=0.84$ for $P=80\%$)

where V is the variance of the case-only estimator of β_3 evaluated under the alternative

Yang et al. (1999) give a sample size formula for the case-only estimator but the variance is given under the null rather than the alternative; we will instead give the variance under the alternative

Sample Size for Case-Only Interactions

$N=(Z_{1-\alpha/2} + Z_{\beta})^2 \times V / \eta^2$

For the case-only estimator, V under the alternative is given by (VanderWeele, 2011; cf. Yang et al., 1999):

$$v_A = (m_1 + m_2 + m_3 + m_4)\left(\frac{1}{m_1} + \frac{1}{m_2} + \frac{1}{m_3} + \frac{1}{m_4}\right)$$

where $m_1 = (1-g)(1-e)$

$m_2 = g(1-e)R_g$

$m_3 = (1-g)eR_e$

$m_4 = geR_gR_eR_i$

Sample Size for Case-Only Interactions

Yang et al. (1999) give a sample size formula for the case-only estimator but the variance is given under the null rather than the alternative and it will again underestimate the same size for large interaction ratios

Nonetheless their comparison with sample sizes for logistic regression and case-control studies is informative

Prevalence of genotype		$R_i^* = 2$ Exposure			
		0.1	0.3	0.5	0.7
0.05	$N_{c\text{-}c}$	4,282	1,929	1,708	2,147
	$N_{c\text{-}o}$	2,996	1,223	997	1,167
0.10	$N_{c\text{-}c}$	2,293	1,041	926	1,167
	$N_{c\text{-}o}$	1,568	659	552	662
0.30	$N_{c\text{-}c}$		486	442	569
	$N_{c\text{-}o}$		308	284	372
0.50	$N_{c\text{-}c}$			414	549
	$N_{c\text{-}o}$			284	400
0.70	$N_{c\text{-}c}$				757
	$N_{c\text{-}o}$				600

This is for α=5% Power=80%

$R_g=R_e=1$, $\exp(\eta)=2$