Clinical Research Methods: Chapter 6

Estimating Sample Size and Power: Applications and Examples

Dr. Alexander Weber

Assistant Professor, Department of Pediatrics,
Division of Neurology, Faculty of Medicine
Imaging Staff Scientist, BC Children's Hospital Research Institute
University of British Columbia

November 17, 2020

Outline

Sample Size Estimation Recipe

1. State **null hypothesis** and whether **one- or two sided alternative hypothesis**
2. Select appropriate **statistical test** based on:
   - type of predictor variable
   - outcome variable
3. Choose **effect size** (and **variability,** if necessary)
4. Set α and β
5. Use appropriate table or formula

Important to estimate sample size early in the design phase

Simple Statistical Tests

TABLE 6.1 Simple Statistical Tests for Use in Estimating Sample Size*

	Outcome Variable	
Predictor Variable	**Dichotomous**	**Continuous**
Dichotomous	Chi-squared test†	*t* test
Continuous	*t* test	Correlation coefficient

* See text for what to do about ordinal variables, or if planning to analyze the data with another type of statistical test.

† The chi-squared test is always two-sided; a one-sided equivalent is the Z statistic.

t Test

t Test

Determine whether the mean value of a continuous outcome variable in one group differs significantly from that in another group

Example

Comparing the mean depression scores in patients treated with an antidepressant and placebo

Assumptions

Distribution of the variable in each of the two groups approximates a normal (bell-shaped) curve
However, the test is remarkably robust (as long as the number of subjects is more than 30; and there are no extreme outliers)

Standardized Effect Size

- Estimate effect size (E) as the difference in the mean value of the outcome variable between the study groups
- Estimate the variability as its standard deviation (S)
- Calculate the standardized effect size: $E \div S$

E and S can often be estimated from previous studies.

Occasionally, a small pilot study will be necessary.

Larger the standardized effect size, the smaller the required sample size.

Standardized effect sizes smaller than 0.1 are difficult to detect and usually not important clinically.

Example Sample Size Calculation with the t Test

Example 6.1 Calculating Sample Size When Using the t Test

Problem: The research question is whether there is a difference in the efficacy of salbutamol and ipratropium bromide for the treatment of asthma. The investigator plans a randomized trial of the effect of these drugs on FEV_1 *(forced expiratory volume in 1 second) after 2 weeks of treatment. A previous study has reported that the mean* FEV_1 *in persons with treated asthma was 2.0 liters, with a standard deviation of 1.0 liter. The investigator would like to be able to detect a difference of 10% or more in mean* FEV_1 *between the two treatment groups. How many patients are required in each group (salbutamol and ipratropium) at* α *(two-sided)* $= 0.05$ *and power* $= 0.80$*?*

Example Sample Size Calculation with the t Test

- E = 0.2 L (10% × 2.0 L)
- S = 1.0 L
- $E \div S = 0.2$ L $\div 1.0$ L $= 0.2$
- $\alpha = 0.05$; $\beta = 1 - 0.8 = 0.2$

TABLE 6A Sample Size *per Group* for Comparing Two Means

One-sided α =		**0.005**			**0.025**			**0.05**	
Two-sided α =		**0.01**			**0.05**			**0.10**	
E/S*	**β =**	**0.05**	**0.10**	**0.20**	**0.05**	**0.10**	**0.20**	**0.05**	**0.10**	**0.20**
0.10		3,565	2,978	2,338	2,600	2,103	1,571	2,166	1,714	1,238
0.15		1,586	1,325	1,040	1,157	935	699	963	762	551
0.20		893	746	586	651	527	394	542	429	310
0.25		572	478	376	417	338	253	347	275	199
0.30		398	333	262	290	235	176	242	191	139
0.40		225	188	148	164	133	100	136	108	78
0.50		145	121	96	105	86	64	88	70	51
0.60		101	85	67	74	60	45	61	49	36
0.70		75	63	50	55	44	34	45	36	26
0.80		58	49	39	42	34	26	35	28	21
0.90		46	39	21	34	27	21	28	22	16
1.00		38	32	26	27	23	17	23	18	14

Shortcut for t Test

- More than 30 subjects
- $\beta = 0.2$
- $\alpha = 0.05$

$$\text{Sample size (per equal-sized group)} = 16 \div (\text{standardized effect size})^2$$

From last example: $16 \div 0.2^2 = 400$ **per group**

The Chi-Squared Test χ^2

χ^2 Test

Compare proportion of subjects in each of two groups who have a dichotomous outcome.

Example

Proportion of men who develop coronary heart disease (CHD) while being treated with folate can be compared with the proportion who develop CHD while taking a placebo.

Assumptions

The χ^2 test is always two sided; an equivalent test for one-sided hypotheses is the **one-sided Z test**

Example Sample Size Calculation with the Chi-Squared Test

Example 6.2 Calculating Sample Size When Using the Chi-Squared Test

Problem: The research question is whether elderly smokers have a greater incidence of skin cancer than nonsmokers. A review of previous literature suggests that the 5-year incidence of skin cancer is about 0.20 in elderly nonsmokers. At α (two-sided) $= 0.05$ and power $= 0.80$, how many smokers and nonsmokers will need to be studied to determine whether the 5-year skin cancer incidence is at least 0.30 in smokers?

Example Sample Size Calculation with the Chi-Squared Test

TABLE 6B.1 Sample Size *per Group* for Comparing Two Proportions

Upper number: $\alpha = 0.05$ (one-sided) or $\alpha = 0.10$ (two-sided); $\beta = 0.20$
Middle number: $\alpha = 0.025$ (one-sided) or $\alpha = 0.05$ (two-sided); $\beta = 0.20$
Lower number: $\alpha = 0.025$ (one-sided) or $\alpha = 0.05$ (two-sided); $\beta = 0.10$

Smaller of P_1 and P_2*	Difference Between P_1 and P_2									
	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
0.05	381	129	72	47	35	27	22	18	15	13
	473	159	88	59	43	33	26	22	18	16
	620	207	113	75	54	41	33	27	23	19
0.10	578	175	91	58	41	31	24	20	16	14
	724	219	112	72	51	37	29	24	20	17
	958	286	146	92	65	48	37	30	25	21
0.15	751	217	108	67	46	34	26	21	17	15
	944	270	133	82	57	41	32	26	21	18
	1,252	354	174	106	73	53	42	33	26	22
0.20	900	251	121	74	50	36	28	22	18	15
	1,133	313	151	91	62	44	34	27	22	18
	1,504	412	197	118	80	57	44	34	27	23
0.25	1,024	278	132	79	53	38	29	23	18	15
	1,289	348	165	98	66	47	35	28	22	18
	1,714	459	216	127	85	60	46	35	28	23
0.30	1,123	300	141	83	55	39	29	23	18	15
	1,415	376	175	103	68	48	36	28	22	18
	1,883	496	230	134	88	62	47	36	28	23
0.35	1,197	315	146	85	56	39	29	23	18	15
	1,509	395	182	106	69	48	36	28	22	18
	2,009	522	239	138	90	62	47	35	27	22

The Correlation Coefficient r

The Correlation Coefficient r

Measure of the strength of the *linear* association between two variables.
It varies between -1 and $+1$;
-1: as one goes up, the other goes down;
$+1$: as one goes up the other goes up;
0: no association

Example

Height and weight in adults are highly correlated in some populations, with $r \approx 0.9$

Warnings

r little intuitive meaning
r^2 is the percentage of variance explained by the relationship between the predictor and outcome variable

Example Sample Size Calculation with the r

Example 6.3 Calculating Sample Size When Using the Correlation Coefficient in a Cross-Sectional Study

Problem: The research question is whether urinary cotinine levels (a measure of the intensity of current cigarette smoking) are correlated with bone density in smokers. A previous study found a modest correlation ($r = -0.3$) between reported smoking (in cigarettes per day) and bone density; the investigator anticipates that urinary cotinine levels will be at least as well correlated. How many smokers will need to be enrolled, at α (two-sided) $= 0.05$ and $\beta = 0.10$?

Example Sample Size Calculation with the r

TABLE 6C Sample Size for Determining Whether a Correlation Coefficient Differs from Zero

One-sided α =	0.005			0.025			0.05		
Two-sided α =	0.01			0.05			0.0101		
β =	0.05	0.10	0.20	0.05	0.10	0.20	0.05	0.10	0.20
r^*									
0.05	7,118	5,947	4,663	5,193	4,200	3,134	4,325	3,424	2,469
0.10	1,773	1,481	1,162	1,294	1,047	782	1,078	854	616
0.15	783	655	514	572	463	346	477	378	273
0.20	436	365	287	319	259	194	266	211	153
0.25	276	231	182	202	164	123	169	134	98
0.30	189	158	125	139	113	85	116	92	67
0.35	136	114	90	100	82	62	84	67	49
0.40	102	86	68	75	62	47	63	51	37
0.45	79	66	53	58	48	36	49	39	29
0.50	62	52	42	46	38	29	39	31	23
0.60	40	34	27	30	25	19	26	21	16
0.70	27	23	19	20	17	13	17	14	11
0.80	18	15	13	14	12	9	12	10	8

* To estimate the total sample size, read across from r (the expected correlation coefficient) and down from the specified values of α and β.

Other Considerations and Special Issues

- Dropouts: after enrolling your sample size, you should expect some people to drop out or not be able to complete the study. In which case, you should increase your recruitment to an even higher number
- Categorical Variables: ordinal variables can often be treated as continuous variables (especially if the number of categories is > 6, and if averaging the values makes sense;
  In other situations, the best strategy may be to dichotomize the categorical variable (e.g. no visits vs. one or more visits per year)
- Survival Analysis: see pg 71
- Clustered Samples: see Chapter 11
- Matching: see Chapter 9
- Multivariate Adjustment: see Chapter 11
- Equivalence Studies: see pg 73

Sample Size Techniques for Descriptive Studies

- Descriptive studies do no have predictor or outcome variables, nor do they compare different groups
- Therefore power and null hypotheses do not apply
- Instead, we calculate descriptive statistics, such as means and proportions

Confidence Intervals

A measure of the precision of a sample estimate
An interval with a greater confidence level (say 99%) is wider, and therefore more likely to include the true population value, than an interval with a lower confidence level (90%).

Sample Size for Confidence Intervals

- The width of a confidence interval depends on the sample size

Example

Estimate the mean score on the U.S. Medical Licensing Examination in a group of medical students
From a sample of 200 students, one might estimate mean score in the population of all students is 215, with a 95% confidence interval from 210 to 220. A smaller study, however, with 50 students, might have about the same mean score but would almost certainly have a wider 95% confidence interval.

- When estimating sample size for descriptive studies, the investigator specifies the desired level and width of the confidence interval. The sample size can then be determined from the tables or formulas in the appendix.

Sample Size for Confidence Intervals Example

Example 6.4 Calculating Sample Size for a Descriptive Study of a Continuous Variable

Problem: The investigator seeks to determine the mean IQ among third graders in an urban area with a 99% confidence interval of ±3 points. A previous study found that the standard deviation of IQ in a similar city was 15 points.

Sample Size for Confidence Intervals Example

TABLE 6D Sample Size for Common Values of W/S*

	Confidence Level		
W/S	90%	95%	99%
0.10	1,083	1,537	2,665
0.15	482	683	1,180
0.20	271	385	664
0.25	174	246	425
0.30	121	171	295
0.35	89	126	217
0.40	68	97	166
0.50	44	62	107
0.60	31	43	74
0.70	23	32	55
0.80	17	25	42
0.90	14	19	33
1.00	11	16	27

* W/S is the standardized width of the confidence interval, computed as W (desired total width) divided by S (standard deviation of the variable). To estimate the total sample size, read across from the *standardized width* and down from the specified confidence level.

Sample Size for Confidence Intervals: Dichotomous Values

- With dichotomous values, results can be expressed as a confidence interval around the estimated proportion of subjects with one of the values

Example 6.5 Calculating Sample Size for a Descriptive Study of a Dichotomous Variable

Problem: The investigator wishes to determine the sensitivity of a new diagnostic test for pancreatic cancer. Based on a pilot study, she expects that 80% of patients with pancreatic cancer will have positive tests. How many such patients will be required to estimate a 95% confidence interval for the test's sensitivity of 0.80 ± 0.05?

Sample Size for Confidence Intervals Example

TABLE 6E Sample Size for Proportions

Upper number: 90% confidence level
Middle number: 95% confidence level
Lower number: 99% confidence level

	Total Width of Confidence Interval (W)						
Expected Proportion (P)*	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.35**	**0.40**
0.10	98	44	—	—	—	—	—
	138	61	—	—	—	—	—
	239	106	—	—	—	—	—
0.15	139	62	35	22	—	—	—
	196	87	49	31	—	—	—
	339	151	85	54	—	—	—
0.20	174	77	44	28	19	14	—
	246	109	61	39	27	20	—
	426	189	107	68	47	35	—
0.25	204	91	51	33	23	17	13
	288	128	72	46	32	24	18
	499	222	125	80	55	41	31
0.30	229	102	57	37	25	19	14
	323	143	81	52	36	26	20
	559	249	140	89	62	46	35
0.40	261	116	65	42	29	21	16
	369	164	92	59	41	30	23
	639	284	160	102	71	52	40
0.50	272	121	68	44	30	22	17

Fixed Sample Size

- Work backwards from sample size
- What effect size can you expect to measure between two groups with a known sample size? (for example)

Example 6.6 Calculating the Detectable Effect Size When Sample Size is Fixed

Problem: An investigator determines that there are 100 patients with systemic lupus erythematosus (SLE) who might be willing to participate in a study of whether a 6-week meditation program affects disease activity, as compared with a control group that receives a pamphlet describing relaxation. If the standard deviation of the change in a validated SLE disease activity scale score is expected to be five points in both the control and the treatment groups, what size difference will the investigator be able to detect between the two groups, at α *(two-sided)* $= 0.05$ *and* $\beta = 0.20$*?*

Strategies for Minimizing Sample Size and Maximizing Power

- Always use continuous when possible

Example 6.7 Use of Continuous versus Dichotomous Variables

Problem: Consider a placebo-controlled trial to determine the effect of nutrition supplements on strength in elderly nursing home residents. Previous studies have established that quadriceps strength (as peak torque in newton-meters) is approximately normally distributed, with a mean of 33 N·m and a standard deviation of 10 N·m, and that about 10% of the elderly have very weak muscles (strength <20 N·m). Nutrition supplements for 6 months are anticipated to increase strength by 5 N·m as compared with the usual diet. This change in mean strength can be estimated, based on the distribution of quadriceps strength in the elderly, to correspond to a reduction in the proportion of the elderly who are very weak to about 5%.

One design might treat strength as a dichotomous variable: very weak versus not very weak. Another might use all the information contained in the measurement and treat strength as a continuous variable. How many subjects would each design require at α (two-sided) = 0.05 and β = 0.20? How does the change in design affect the research question?

Solution with Dichotomous Values

- Alternative Hypothesis: Proportion of elderly residents who are very weak after receiving 6 months of nutrition supplements differs from the proportion in those on a usual diet
- Prevalence of being very weak on usual diet = 0.10
- Prevalence of being very weak if the vitamins work = 0.05

TABLE 6B.1 Sample Size *per Group* for Comparing Two Proportions

Upper number: $\alpha = 0.05$ (one-sided) or $\alpha = 0.10$ (two-sided); $\beta = 0.20$
Middle number: $\alpha = 0.025$ (one-sided) or $\alpha = 0.05$ (two-sided); $\beta = 0.20$
Lower number: $\alpha = 0.025$ (one-sided) or $\alpha = 0.05$ (two-sided); $\beta = 0.10$

Smaller of P_1 and P_2*	Difference Between P_1 and P_2									
	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
0.05	381	129	72	47	35	27	22	18	15	13
	473	159	88	59	43	33	26	22	18	16
	620	207	113	75	54	41	33	27	23	19
0.10	578	175	91	58	41	31	24	20	16	14
	724	219	112	72	51	37	29	24	20	17
	958	286	146	92	65	48	37	30	25	21
0.15	751	217	108	67	46	34	26	21	17	15
	944	270	133	82	57	41	32	26	21	18
	1,252	354	174	106	73	53	42	33	26	22
0.20	900	251	121	74	50	36	28	22	18	15
	1,133	313	151	91	62	44	34	27	22	18
	1,504	412	197	118	80	57	44	34	27	23
0.25	1,024	278	132	79	53	38	29	23	18	15
	1,289	348	165	98	66	47	35	28	22	18
	1,714	459	216	127	85	60	46	35	28	23

Solution with Continuous Values

TABLE 6A Sample Size *per Group* for Comparing Two Means

One-sided α =	0.005			0.025			0.05		
Two-sided α =	0.01			0.05			0.10		
*E/S** β =	0.05	0.10	0.20	0.05	0.10	0.20	0.05	0.10	0.20
0.10	3,565	2,978	2,338	2,600	2,103	1,571	2,166	1,714	1,238
0.15	1,586	1,325	1,040	1,157	935	699	963	762	551
0.20	893	746	586	651	527	394	542	429	310
0.25	572	478	376	417	338	253	347	275	199
0.30	398	333	262	290	235	176	242	191	139
0.40	225	188	148	164	133	100	136	108	78
0.50	145	121	96	105	86	64	88	70	51
0.60	101	85	67	74	60	45	61	49	36
0.70	75	63	50	55	44	34	45	36	26
0.80	58	49	39	42	34	26	35	28	21
0.90	46	39	21	34	27	21	28	22	16
1.00	38	32	26	27	23	17	23	18	14

* E/S is the standardized effect size, computed as E (expected effect size) divided by S (SD of the outcome variable). To estimate the sample size, read across from the *standardized effect size*, and down from the specified values of α and β for the required sample size in each group.

Solution with Continuous Values

TABLE 6A Sample Size *per Group* for Comparing Two Means

One-sided α =	0.005			0.025			0.05		
Two-sided α =	0.01			0.05			0.10		
*E/S** β =	0.05	0.10	0.20	0.05	0.10	0.20	0.05	0.10	0.20
0.10	3,565	2,978	2,338	2,600	2,103	1,571	2,166	1,714	1,238
0.15	1,586	1,325	1,040	1,157	935	699	963	762	551
0.20	893	746	586	651	527	394	542	429	310
0.25	572	478	376	417	338	253	347	275	199
0.30	398	333	262	290	235	176	242	191	139
0.40	225	188	148	164	133	100	136	108	78
0.50	145	121	96	105	86	64	88	70	51
0.60	101	85	67	74	60	45	61	49	36
0.70	75	63	50	55	44	34	45	36	26
0.80	58	49	39	42	34	26	35	28	21
0.90	46	39	21	34	27	21	28	22	16
1.00	38	32	26	27	23	17	23	18	14

* E/S is the standardized effect size, computed as E (expected effect size) divided by S (SD of the outcome variable). To estimate the sample size, read across from the *standardized effect size*, and down from the specified values of α and β for the required sample size in each group.

- Using dichotomous values: 473 *per group*
- Using continuous values: 64 *per group*

Use Paired Measurements

Paired Measurements

Longitudinal study: compare before and after in same subjects

- Requires smaller sample size because, by comparing each subject with herself, it removes the baseline between-subject part of the variability of the outcome variable.
- Can be two-sampled or one-sampled
- Two-sampled: comparing the mean values of an outcome variable in two groups in a pair of measurements, say before and after an intervention
- One-sampled: compares the mean change in a pair of measurements within a single group to zero change
- Caution: one-sampled paired t test is a fairly weak design due the absence of a comparison group; makes it difficult to know what would have happened had the subjects been left untreated

Paired t Test Example

Example 6.8 Use of the t Test with Paired Measurements

Problem: Recall Example 6.1, in which the investigator studying the treatment of asthma is interested in determining whether salbutamol can improve FEV_1 by 200 mL compared with ipratropium bromide. Sample size calculations indicated that 394 subjects per group are needed, more than are likely to be available. Fortunately, a colleague points out that asthmatic patients have great differences in their FEV_1 values before treatment. These between-subject differences account for much of the variability in FEV_1 after treatment, therefore obscuring the effect of treatment. She suggests using a paired t test to compare the changes in FEV_1 in the two groups. A pilot study finds that the standard deviation of the change in FEV_1 is only 250 mL. How many subjects would be required per group, at α (two-sided) $= 0.05$ and $\beta = 0.20$?

Paired t Test Example

TABLE 6A Sample Size *per Group* for Comparing Two Means

One-sided α =		0.005			0.025			0.05	
Two-sided α =		0.01			0.05			0.10	
*E/S**	β = 0.05	0.10	0.20	0.05	0.10	0.20	0.05	0.10	0.20
0.10	3,565	2,978	2,338	2,600	2,103	1,571	2,166	1,714	1,238
0.15	1,586	1,325	1,040	1,157	935	699	963	762	551
0.20	893	746	586	651	527	394	542	429	310
0.25	572	478	376	417	338	253	347	275	199
0.30	398	333	262	290	235	176	242	191	139
0.40	225	188	148	164	133	100	136	108	78
0.50	145	121	96	105	86	64	88	70	51
0.60	101	85	67	74	60	45	61	49	36
0.70	75	63	50	55	44	34	45	36	26
0.80	58	49	39	42	34	26	35	28	21
0.90	46	39	21	34	27	21	28	22	16
1.00	38	32	26	27	23	17	23	18	14

* E/S is the standardized effect size, computed as E (expected effect size) divided by S (SD of the outcome variable). To estimate the sample size, read across from the *standardized effect size*, and down from the specified values of α and β for the required sample size in each group.

Use More Precise Variables

- In a t test, for example, a 20% decrease in SD of the outcome variable results in a 36% decrease in the sample size
- Strategies for Enhancing Precision (from Chpt 4):
  1. Standardize the measurement method
  2. Train and certify the observers
  3. Refine the instrument
  4. Automate the instrument
  5. Repeating the measurement, and taking the mean of the two or more readings

Use Unequal Group Sizes

- Usually it's easier to recruit more controls than patients
- In general, the gain in power when the size of one group is increased to twice the size of the other is considerable
- There are diminishing returns in triple or quadruple sizes
- Useful approximation for: case-control studies of dichotomous risk factors and outcomes using c controls per case:
  $n' = [(c + 1) \div 2c] \times n$

Example 6.9 Use of Multiple Controls per Case in a Case–Control Study

Problem: An investigator is studying whether exposure to household insecticide is a risk factor for aplastic anemia. The original sample size calculation indicated that 25 cases would be required, using one control per case. Suppose that the investigator has access to only 18 cases. How should the investigator proceed?

How to Estimate Sample Size When There is Insufficient Information?

- Extensive search: roughly comparable situations and mediocre or dated findings may be good enough
- Pilot study: highly recommended for almost all studies that involve new instruments, measurement methods, or recruitment strategies
- For continuous variables with roughly bell-shaped distribution: SD can be estimated as one-quarter of the difference between the high and low ends of the range of values that occur commonly
  - Example: if most subjects are likely to have a serum sodium level between 135 and 143 mEq/L, the SD of serum sodium is about 2 mEq/L (0.25×8 mEq/L
- Effect size can be calculated from what is believed to be clinically meaningful

Common Errors to Avoid

1. Estimate sample size early!
2. Dichotomous variables can appear to be continuous when they are expressed as a percentage or rate; however, these are still dichotomous, and a chi-squared test should be used
3. Sample size represents the # of subjects needed with outcome data, not the number to be enrolled: investigator should plan for dropouts and missing data

Summary

- When estimating sample size for an analytic study:
  1. state null and alternative hypothesis; state how many sides
  2. pick statistical test based on types of predictor and outcome variables
  3. estimate effect size (and variability)
  4. specify appropriate values for α and β, based on importance of avoiding Type I and II errors
- Adjust for potential dropouts
- When estimating sample size for descriptive studies:
  1. estimate the proportion of subjects with a dichotomous outcome or the standard deviation of a continuous outcome
  2. specify the desired precision (width of the confidence interval)
  3. specify the confidence level (e.g., 95%)

Summary

- When sample size is predetermined, the investigator can work backward to estimate the detectable effect size or, less commonly, the power
- Strategies to minimize the required sample size include:
  1. using continuous variables,
  2. more precise measurements,
  3. paired measurements,
  4. unequal group sizes,
  5. and more common outcomes.
- When there seems to not be enough information to estimate the sample size, the investigator should review the literature in related areas, do a small pilot study or choose an effect size that is clinically meaningful; standard deviation can be estimated as 1/4 of the range of commonly encountered values. If none of these is feasible, an educated guess can give a useful ballpark estimate.