Download

http://www.palmx.org/samplesize/ samplesize-2009.pdf

HOW TO CALCULATE YOUR OWN SAMPLE SIZE

or Please don't disturb us in JKM anymore...

The Magic Number

How many subjects do I need to obtain a significant result for my study?
In medical research, the sample size has to be "just large enough".
If too small, it's a waste of time doing the study since no conclusive results are likely to be obtained.

What happens if the sample size is too small?

Data of a clinical trial on 30 patients on comparison of pain control between two modes of treatment.

Type of treatment * Pain (2 hrs post-op) Crosstabulation						
			Pain (2 hrs			
			No pain	In pain	Total	
Type of treatment	Pethidine	Count	<u>8</u>	_ 7_	15	
		% within Type of treatment	53.3%	46.7%	100.0%	
	Cocktail	Count	4	11	15	
		% within Type of treatment	26.7%	73.3%	100.0%	
Total		Count	12	18	30	
		% within Type of treatment	40.0%	60.0%	100.0%	

Chi-square =2.222, p=0.136

p = 0.136. p bigger than 0.05. No significant difference and the null hypothesis was not rejected.

There was a large difference between the rates (53.3% vs 26.7%) but the result was not significant.

© Dr Azmi Mohd Tamil, 2009 Not signi the stuc	ficant sinc ly is less t	e power of nan 80%.
😿 Power and Sample Size Program: Main Win	dow 📃 🗖 🔁	3
File Log Help		
Output	egression 2 Dichotomous Log	
What do you want to know? Power Power for uncorrected chi-squared test .3208		Power is only
Design Matched or Independent?	Independent	32%!
Case control?	Prospective	
How is the alternative hypothesis expressed? Uncorrected chi-square or Fisher's exact test?	Two proportions	
Input	Graphs	
Logging is enabled.	Exit	

© Dr Azmi Mohd Tamil, 2009 **For power of the study of 80%, sample size required is;**

😿 Power and Sample Size Program: Main Window				
File Log Help				
Survival t-test Regression 1 Regression 2 Dichotor	nous Log			
Studies that are analysed by chi-square or Fi Output What do you want to know? Sample size	sher's exact test			
Case sample size for uncorrected 50 <u>chi-squared test</u> Design				
Matched or Independent? Independent				
Case control? Prospective				
How is the alternative hypothesis expressed? Two proportions				
Uncorrected chi-square or Fisher's exact test? Uncorrected chi-square test				
Input $\underline{\alpha} \stackrel{.05}{=} \begin{array}{c} \underline{p}_{0} \stackrel{.53}{=} \\ \underline{power} \stackrel{.8}{=} \begin{array}{c} \underline{p}_{1} \stackrel{.26}{=} \\ \underline{p}_{1} \stackrel{.26}{=} \end{array}$	Graphs			
Logging is enabled.	Exit			

Sample size reqd. 50 per group. For case & control, total 100!

Inadequate sample size



By increasing the sample size.....

Data of a clinical trial on 90 (instead of 30!) patients on comparison of pain control between two modes of treatment.

Type of treatment * Pain (2 hrs post-op) Crosstabulation						
			Pain (2 hr			
			No pain	In pain	Total	
Type of treatment	Pethidine	Count	24	21	45	
		% within Type of treatment	53.3%	46.7%	100.0%	
	Cocktail	Count	12	33	45	
		% within Type of treatment	26.7%	73.3%	100.0%	
Total		Count	36	54	90	
		% within Type of treatment	40.0%	60.0%	100.0%	

Chi-square =6.667, p=0.01

Now p = 0.01. p smaller than 0.05. There was significant difference and the null hypothesis was rejected.

The difference between the rates (53.3% vs 26.7%) are the same but the result was significant with the larger sample size.

Conclusion

Same difference of rate but only significant with larger sample size (n=90 instead of 30). Therefore the sample size has to be "just large enough". If too small, it's a waste of time doing the study since no conclusive results are likely to be obtained.

Why do we calculate the required sample size?

- Cost of the study based on the sample size required more sample, higher cost
 - Estimate the length of the study i.e if sample needed is 120 patients and each year only 40 patients available, need at least 3 years to complete.
- Feasible or not? Whether within the constraints of time allocated (i.e. 1 year) and budget available (i.e. RM5000).

Why do you calculate it? So that your research proposal is approved by the ethical committee. ;-)

What is power?

It is the probability of finding an effect given that one truly exists

p = probability of observing this data given that H₀ true

Power (denoted $1-\beta$) = probability of finding p<alpha given that H_0 false

and thus to be reasonably sure that no such benefit exists, if it is not found in the trial.

	Powe	er and a Real	pha world
	H ₀ Not Rejected	H ₀ True True negative	H ₀ False Type II error ('miss')
Experimental result		Probability = 1- α	Probability = β
	H ₀ Rejected	Type I error ('false alarm') Probability = α	True positive ('hit') Probability = 1-β

INTRODUCTION

- The greater the power of study, the more sure we can be, but greater power requires a larger sample.
 - It is common to require a power of between 80% and 90%.
- Power of 80% for detecting effect difference
- Power of 90% for proving equal effect (equivocal studies)

Power is affected by sample size





INTRODUCTION

- Sample size calculations are based on the quantity known as the effect size.
- The smaller the effect size, the larger the required size of the sample.
 - For example, if treated group improved 10x better than the control group, the sample size required is only 242.
- But if the treated group improved only 5x better than the control group, the sample size required is 664.

Effect Size

- 'Effect Size' is simply a way of quantifying the difference between two groups.
 - For example, if one group has had an 'experimental' treatment and the other has not (the 'control'), then the Effect Size is a measure of the effectiveness of the treatment.

Different Types of Effect Sizes:

Standardised Mean Difference

- Comparing group research
 - Either treatment groups
 - Or naturally occurring groups
- inherently continuous construct (continuous outcome-mean)

Odds-Ratio

- Comparing group research
 - Either treatment groups
 - Or naturally occurring groups
- inherently dichotomous construct (categorical outcome-rate)
- Correlation Coefficient
 - association between variables research

Different Types of Effect Sizes:

Proportion

- Measures of central tendency research
 - HIV/AIDS prevalence rates
 - Proportion of homeless persons found to be alcohol abusers

Standardised Gain Score

- gain or change between two measurement points on the same variable
 - reading speed before and after a reading improvement class

Effect size formula

Effect _ size = $\frac{\mu_0 - \mu_1}{\mu_0 - \mu_1}$ $\boldsymbol{\sigma}$

where σ is standard deviation of population of dependent (outcome) measure scores.

Cohen's Effect Size (d)

Cohen (1992) gives the following guidelines for the social sciences:
small effect size, 0.2;
medium, 0.5;
large, 0.8.

Factors governing power

Power, $1-\beta$ = probability of finding an effect, given that there actually is one

So power will obviously be governed by

- Effect size (stronger effect size, more power)
- Number of subjects (more subjects, more power)
- Choice of alpha (0.01 need more, 0.05 need less)
- Also (maybe less obviously)...
 - Sources of variability (i.e. sampling method)
 - Study design (case-control vs cohort vs clinical trial)
 - Choice of statistical test (chi² or t-test)

Calculating power & sample size

- As you've just seen, power is determined by
 - Effect Size
 - Number of subjects
 - Choice of alpha (usually 0.05)

SPSS doesn't have the facility to calculate power/sample size but you can download a free program (PS2) to do so (given the above information) from

http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/PowerSampleSize

Or you can use Statcalc (part of EpiInfo6) from <u>http://www.cdc.gov/epiinfo/Epi6/El6dnjp.htm</u> to calculate sample size

PS for Power/Sample Size calculations

	😿 Power and Sample Size Program: Main Window					
	File Log Help					
	Survival	t-test	Regression 1	Regression 2	Dichotomous	Log
	Output			<u>Studies</u>	that are analyse	ed by t-tests
	What d	<u>o you want to kno</u>	W? Powe	er		•
	Power		.7989			
	Design	- index and - 10		Paire	ad .	_
Any blue label leads to help		or independent?		Ji dire		
	Input	<u>x</u> 005	<u>ð</u> 70	\supset	[Calculate
Delta means expected difference in scores		<u>n</u> 25	<u>a</u> 12			Graphs
Expected standard						
differences	Logging is en	abled.				Exit

^{© Dr Azm} PS Tor Power/Sample Size calculations

- Alternatively you can ask the program to calculate the sample size you need to give the *power* that you're looking for
- Also plots graphs for a range of powers etc...



25

PS2 can calculate sample size for

- Independent Case-Control Studies: Chi-square test, Fisher's exact test.
- Matched Case-Control Studies: McNemar's Test.
 - Cohort Studies With Dichotomous Outcomes: Chi-square test, McNemar's test
- Continuous Response Measures in Two Groups: Paired and independent t tests.
- Linear Regression
- Survival Studies

StatCalc can calculate sample size Cross-Sectional – Prevalence Studies Cross-Sectional – Categorical Risk Factor & Outcome **Cohort Studies With Dichotomous Outcomes: Categorical Risk Factor &** Outcome Unmatched Case-Control Studies: Categorical Risk Factor & Outcome

Or refer to tables;

- Lwanga SK, Lemeshow S., 1991. Sample Size Determination in Health Studies: A Practical Manual. WHO.
- A pdf copy of the book is available.
- Only useful for prevalence studies.

SAMPLE SIZE DETERMINATION IN HEALTH STUDIES

0'

A Practical Manual

S. K. Lwanga Epidemiological and Statistical Methodology World Health Organization Geneva, Switzeriand And S. Lemeshow Division of Public Health University of Massachusetts af Amfrest

1991

World Health Organizatio

Other Sample Size Estimation Software

License software

nQuery Advisor version 6.0.

Free online calculator

- http://www.changbioscience.com/stat/ssize.ht ml
- http://calculators.stat.ucla.edu/

Step One

Determine your research design and the outcome being measured

STUDY DESIGNS



Examples of Study Design

- Study the prevalence of obesity in HUKM crosssectional (prevalence)
- Comparing the rate of diabetes mellitus between Indians and non-Indians – cross-sectional (comparative)
- Comparing the rate of diabetes mellitus between those with cataract and those without cataract case-control
- Measure the incidence and relative risk of diabetes mellitus between normal and overweight - cohort
- Comparing Fluoxetine against Sertraline for treating depression – clinical trial
- Measuring the sensitivity and specificity of a new serological test against the gold standard –

diagnostic study

Step 2 – Go to the Respective Slides for that Design

Cross Sectional Study (prevalence) – Go to 34
Cross Sectional Study (comparative) – Go to 50
Case Control – Go to 62
Cohort – Go to 70
Clinical Trial (qualitative outcome) – Go to 78
Clinical Trial (continuous outcome) – Go to 90
Diagnostic Study – Go to 101
Each design requires a different approach for sample size calculation

Calculate Your Own Sample Size

Cross-Sectional Study – Measuring Prevalence

Cross-Sectional

What is the outcome being measured?

- Is it the prevalence of disease/risk factor?
- As tween the main ical risk factor
 Ca catcome?

Prevalence in **Cross-Sectional**

 Do a literature review to estimate the prevalence being studied.
 Determine the absolute precision required i.e. 5 percentage points (usually between 3 to 5).
 Calculate using (Kish L. 1965) n = (Z_{1-α})²(P(1-P)/D²)
 or refer to a table in S.K. Lwanga, S. Lemeshaw 1991, Sample Size Determination in Health Studies, pg 25
 Or use StatCalc from Epilnfo6.
Example – Prevalence of Obesity in HUKM

 Confidence interval = 1 - α = 95%; Z_{1-α} = Z_{0.95} = 1.96 (from normal distribution table).
 Prevalence = P = 20%
 Absolute precision required = 5 percentage points,
 (if the colored prevalence of the study is

(if the calculated prevalence of the study is 20%, then the true value of the prevalence lies between 15-25%).

Calculate Manually

n = (Z_{1-α})²(P(1-P)/D²) where
 Z_{1-α} = Z_{0.95} = 1.96 (from normal distribution table. This value of 1.96 is standard for CI of 95%).
 P = 20% = 0.2 in this example
 D = 5% = 0.05 in this example
 n = 1.96² x (0.2(1-0.2)/0.05²) = 245.84

Refer to Table

Refer to the table in S.K. Lwanga, S. Lemeshaw 1991, Sample Size Determination in Health Studies pg 25.
 With a Prevalence (P) of 20%, precision of 0.05, the table indicates that the sample size required is 246.

Table 1. Estimating a population proportion with specified absolute precision

 $n = z_{1-a/2}^2 P(1-P)/d^2$

(a) Confidence level 95%

I.

0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
1825	3457	4898	6147	7203	8067	8740	9220	9508	9604	9508	9220	8740	8067	7203	6147	4898	3457	1825
456	864	1225	1537	1801	2017	2185	2305	2377	2401	2377	2305	2185	2017	1801	1537	1225	864	456
203	384	544	683	800	896	971	1024	1056	1067	1056	1024	971	896	800	683	544	384	203
114	216	306	384	450	504	546	576	594	600	594	576	546	504	450	384	306	216	114
73	138	196	246	28 8	323	350	369	380	384	380	369	350	323	288	246	196	138	73
51	96	136	171	200	224	243	256	264	267	264	256	243	224	200	171	136	96	st. 51/
37	- 71	100	125	147	165	178	188	194	196	194	188	178	165	147	125	100	71	37
29	54	, 77	96	113	126	137	144	149	150	149	144	137	126	113	96	77	54	29
23	43	60	76	89	100	108	114	117	119	117	114	108	100	89	76	60	43 -	23
18	35	_ 49	61	72	81	87	92	9 5	9,6	95	92	87	81	72	61	49	35	18
15	29	40	51	60	67	72	76	79	79	79	76	72	67	60	51	40	29	15
13	24	34	43	50	56	61	64	66	67	66	64	61	56	50	43	34	24	13
11	20	29	36	43	48	52	55	56	57	56	55	52	48	43	36	29	20	11
9	18	25	31	37	41	45	47	49	49	49	47	45	41	37	31	25	18	9
8	15	22	27	32	36	39	41	42	43	42	41	39	36	32	27	22	15	8
5	9	12	15	18	20	22	23	24	24	24	23	22	20	. 18	15	. 12	9	Б
•	6	8	10	12	13	14	15	15	15	15	15	14	13	12	10	8	6	
	0.05 1825 456 203 114 73 51 37 29 23 18 15 13 11 9 8 5	0.05 0.10 1825 3457 456 864 203 384 114 216 73 138 51 96 37 71 29 54 23 43 18 35 15 29 13 24 11 20 9 18 8 15 5 9 6	0.05 0.10 0.15 1825 3457 4898 456 864 1225 203 384 544 114 216 306 73 138 196 51 96 136 37 71 100 29 54 77 23 43 60 18 35 49 15 29 40 13 24 34 11 20 29 9 18 25 8 15 22 5 9 12 6 8 8	0.05 0.10 0.15 0.20 1825 3457 4898 6147 456 864 1225 1537 203 384 544 683 114 216 306 384 73 138 196 246 51 96 136 171 37 71 100 125 29 54 77 96 23 43 60 76 18 35 49 61 15 29 40 51 13 24 34 43 11 20 29 36 9 18 25 31 8 15 22 27 5 9 12 15 6 8 10	0.05 0.10 0.15 0.20 0.25 1825 3457 4898 6147 7203 456 864 1225 1537 1801 203 384 544 683 800 114 216 306 384 450 73 138 196 246 288 51 96 136 171 200 37 71 100 125 147 29 54 77 96 113 23 43 60 76 89 18 35 49 61 72 15 29 40 51 60 13 24 34 43 50 11 20 29 36 43 9 18 25 31 37 8 15 22 27 32 5 9 12 15 18 6 8 10 12	0.05 0.10 0.15 0.20 0.25 0.30 1825 3457 4898 6147 7203 8067 456 864 1225 1537 1801 2017 203 384 544 683 800 896 114 216 306 384 450 504 73 138 196 246 288 323 51 96 136 171 200 224 37 71 100 125 147 165 29 54 77 96 113 126 23 43 60 76 89 100 18 35 49 61 72 81 15 29 40 51 60 67 13 24 34 43 50 56 11 20 29 36 43 48 9 18 25 31 37 41 8 15 22	0.05 0.10 0.15 0.20 0.25 0.30 0.35 1825 3457 4898 6147 7203 8067 8740 456 864 1225 1537 1801 2017 2185 203 384 544 683 800 896 971 114 216 306 384 450 504 546 73 138 196 246 288 323 350 51 96 136 171 200 224 243 37 71 100 125 147 165 178 29 54 77 96 113 126 137 23 43 60 76 89 100 108 18 35 49 61 72 81 87 15 29 40 51 60 67 72 13 24 34 43 50 56 61 11 20 29 3	0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 1825 3457 4898 6147 7203 8067 8740 9220 456 864 1225 1537 1801 2017 2185 2305 203 384 544 683 800 896 971 1024 114 216 306 384 450 504 546 576 73 138 196 246 288 323 350 369 51 96 136 171 200 224 243 256 37 71 100 125 147 165 178 188 29 54 77 96 113 126 137 144 23 43 60 76 89 100 108 114 18 35 49 61 72 81 87 92 15 29 40 51 60 67 72 76	0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 1825 3457 4898 6147 7203 8067 8740 9220 9508 456 864 1225 1537 1801 2017 2185 2305 2377 203 384 544 683 800 896 971 1024 1056 114 216 306 384 450 504 546 576 594 73 138 196 246 288 323 350 369 380 51 96 136 171 200 224 243 256 264 37 71 100 125 147 165 178 188 194 29 54 77 96 113 126 137 144 149 23 43 60 76 89 100 108 114 117 18 35 49 61 72 81	0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 1825 3457 4898 6147 7203 8067 8740 9220 9508 9604 456 864 1225 1537 1801 2017 2185 2305 2377 2401 203 384 544 683 800 896 971 1024 1056 1067 114 216 306 384 450 504 546 576 594 600 73 138 196 246 288 323 350 369 380 384 51 96 136 171 200 224 243 256 264 267 37 71 100 125 147 165 178 188 194 196 23 43 60 76 89 100 108 114 11	0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 1825 3457 4898 6147 7203 8067 8740 9220 9508 9604 9508 456 864 1225 1537 1801 2017 2185 2305 2377 2401 2377 203 384 544 683 800 896 971 1024 1056 1067 1056 114 216 306 384 450 504 546 576 594 600 594 73 138 196 246 288 323 350 369 380 384 380 51 96 136 171 200 224 243 256 264 267 264 37 71 100 125 147 165 178 188 194 196 194 29 54 77 96 113 126 137 144 149	0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 1825 3457 4898 6147 7203 8067 8740 9220 9508 9604 9508 9220 456 864 1225 1537 1801 2017 2185 2305 2377 2401 2377 2305 203 384 544 683 800 896 971 1024 1056 1067 1056 1024 114 216 306 384 450 504 546 576 594 600 594 576 73 138 196 246 288 323 350 369 380 384 380 369 51 96 136 171 200 224 243 256 264 267 264 256 37 71 100 125 147 165 178 188 194 196 194 188 29 <td>0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 1825 3457 4898 6147 7203 8067 8740 9220 9508 9604 9508 9200 8740 456 864 1225 1537 1801 2017 2185 2305 2377 2401 2377 2305 2185 203 384 544 683 800 896 971 1024 1056 1067 1056 1024 971 114 216 306 384 450 504 546 576 594 600 594 566 243 73 138 196 246 288 323 350 369 380 384 380 369 350 51 96 136 171 200 224 243 256 264 267 264 2</td> <td>0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 1825 3457 4898 6147 7203 8067 8740 9220 9508 9604 9508 9220 8740 8067 456 864 1225 1537 1801 2017 2185 2305 2377 2401 2377 2305 2185 2017 203 384 544 683 800 896 971 1024 1056 1067 1056 1024 971 896 114 216 306 384 450 504 546 576 594 600 594 576 546 504 73 138 196 246 288 323 350 369 380 384 380 369 350 323 51 96 135 177 165 <t< td=""><td>0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 1825 3457 4898 6147 7203 8067 8740 9220 9508 9604 9508 9220 8740 8067 7203 456 864 1225 1537 1801 2017 2185 2305 2377 2401 2377 2305 2185 2017 1801 203 384 544 683 800 896 971 1024 1056 1067 1056 1024 971 896 800 114 216 306 384 450 504 546 576 594 600 594 576 546 504 450 73 138 196 246 288 323 350 369 380 384 380 369 350 323 288</td><td>0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 1825 3457 4898 6147 7203 8067 8740 9220 9508 9604 9508 9220 8740 8067 7203 6147 456 864 1225 1537 1801 2017 2185 2305 2377 2401 2377 2305 2185 2017 1801 1537 203 384 544 683 800 896 971 1024 1056 1067 1056 1024 971 896 800 683 114 216 306 384 450 504 546 576 594 600 594 576 546 504 450 384 73 138 196 246 288 323 350 324 266 243</td><td>0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 1825 3457 4898 6147 7203 8067 8740 9220 9508 9604 9508 9220 8740 8067 7203 6147 4898 456 B64 1225 1537 1801 2017 2185 2305 2377 2401 2377 2305 2185 2017 1801 1537 1225 203 384 544 683 800 896 971 1024 1056 1067 1056 1024 971 896 800 683 544 114 216 306 384 450 504 564 564 504 503 323 288 246 196 51 96 136 171 200 224 243 256<td>0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 1825 3457 4898 6147 7203 8067 8740 9220 9508 9604 9508 9220 8740 8067 7203 6147 4898 3457 456 864 1225 1537 1801 2017 2185 2305 2377 2401 2377 2305 2185 2017 1801 1537 1225 864 103 384 544 683 800 896 971 1024 1056 1067 1056 1024 971 896 800 683 544 384 114 216 306 384 450 504 546 576 594 600 594 576 546 504 450 384 306 216 73 138 196 137 144 195 188 384 380</td></td></t<></td>	0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 1825 3457 4898 6147 7203 8067 8740 9220 9508 9604 9508 9200 8740 456 864 1225 1537 1801 2017 2185 2305 2377 2401 2377 2305 2185 203 384 544 683 800 896 971 1024 1056 1067 1056 1024 971 114 216 306 384 450 504 546 576 594 600 594 566 243 73 138 196 246 288 323 350 369 380 384 380 369 350 51 96 136 171 200 224 243 256 264 267 264 2	0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 1825 3457 4898 6147 7203 8067 8740 9220 9508 9604 9508 9220 8740 8067 456 864 1225 1537 1801 2017 2185 2305 2377 2401 2377 2305 2185 2017 203 384 544 683 800 896 971 1024 1056 1067 1056 1024 971 896 114 216 306 384 450 504 546 576 594 600 594 576 546 504 73 138 196 246 288 323 350 369 380 384 380 369 350 323 51 96 135 177 165 <t< td=""><td>0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 1825 3457 4898 6147 7203 8067 8740 9220 9508 9604 9508 9220 8740 8067 7203 456 864 1225 1537 1801 2017 2185 2305 2377 2401 2377 2305 2185 2017 1801 203 384 544 683 800 896 971 1024 1056 1067 1056 1024 971 896 800 114 216 306 384 450 504 546 576 594 600 594 576 546 504 450 73 138 196 246 288 323 350 369 380 384 380 369 350 323 288</td><td>0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 1825 3457 4898 6147 7203 8067 8740 9220 9508 9604 9508 9220 8740 8067 7203 6147 456 864 1225 1537 1801 2017 2185 2305 2377 2401 2377 2305 2185 2017 1801 1537 203 384 544 683 800 896 971 1024 1056 1067 1056 1024 971 896 800 683 114 216 306 384 450 504 546 576 594 600 594 576 546 504 450 384 73 138 196 246 288 323 350 324 266 243</td><td>0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 1825 3457 4898 6147 7203 8067 8740 9220 9508 9604 9508 9220 8740 8067 7203 6147 4898 456 B64 1225 1537 1801 2017 2185 2305 2377 2401 2377 2305 2185 2017 1801 1537 1225 203 384 544 683 800 896 971 1024 1056 1067 1056 1024 971 896 800 683 544 114 216 306 384 450 504 564 564 504 503 323 288 246 196 51 96 136 171 200 224 243 256<td>0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 1825 3457 4898 6147 7203 8067 8740 9220 9508 9604 9508 9220 8740 8067 7203 6147 4898 3457 456 864 1225 1537 1801 2017 2185 2305 2377 2401 2377 2305 2185 2017 1801 1537 1225 864 103 384 544 683 800 896 971 1024 1056 1067 1056 1024 971 896 800 683 544 384 114 216 306 384 450 504 546 576 594 600 594 576 546 504 450 384 306 216 73 138 196 137 144 195 188 384 380</td></td></t<>	0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 1825 3457 4898 6147 7203 8067 8740 9220 9508 9604 9508 9220 8740 8067 7203 456 864 1225 1537 1801 2017 2185 2305 2377 2401 2377 2305 2185 2017 1801 203 384 544 683 800 896 971 1024 1056 1067 1056 1024 971 896 800 114 216 306 384 450 504 546 576 594 600 594 576 546 504 450 73 138 196 246 288 323 350 369 380 384 380 369 350 323 288	0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 1825 3457 4898 6147 7203 8067 8740 9220 9508 9604 9508 9220 8740 8067 7203 6147 456 864 1225 1537 1801 2017 2185 2305 2377 2401 2377 2305 2185 2017 1801 1537 203 384 544 683 800 896 971 1024 1056 1067 1056 1024 971 896 800 683 114 216 306 384 450 504 546 576 594 600 594 576 546 504 450 384 73 138 196 246 288 323 350 324 266 243	0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 1825 3457 4898 6147 7203 8067 8740 9220 9508 9604 9508 9220 8740 8067 7203 6147 4898 456 B64 1225 1537 1801 2017 2185 2305 2377 2401 2377 2305 2185 2017 1801 1537 1225 203 384 544 683 800 896 971 1024 1056 1067 1056 1024 971 896 800 683 544 114 216 306 384 450 504 564 564 504 503 323 288 246 196 51 96 136 171 200 224 243 256 <td>0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 1825 3457 4898 6147 7203 8067 8740 9220 9508 9604 9508 9220 8740 8067 7203 6147 4898 3457 456 864 1225 1537 1801 2017 2185 2305 2377 2401 2377 2305 2185 2017 1801 1537 1225 864 103 384 544 683 800 896 971 1024 1056 1067 1056 1024 971 896 800 683 544 384 114 216 306 384 450 504 546 576 594 600 594 576 546 504 450 384 306 216 73 138 196 137 144 195 188 384 380</td>	0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 1825 3457 4898 6147 7203 8067 8740 9220 9508 9604 9508 9220 8740 8067 7203 6147 4898 3457 456 864 1225 1537 1801 2017 2185 2305 2377 2401 2377 2305 2185 2017 1801 1537 1225 864 103 384 544 683 800 896 971 1024 1056 1067 1056 1024 971 896 800 683 544 384 114 216 306 384 450 504 546 576 594 600 594 576 546 504 450 384 306 216 73 138 196 137 144 195 188 384 380

.

'Sample size less than 5.

Alternative to table

http://www.palmx.org/samplesize/Calc_Samplesize.xls

	n =	(Z _{1-α})2 Prev	(P(1-P	9)/d²) e (P)		Conf ©dr	idence tamil@	e level ⊋medi	95% c. ukm	.my										
		0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
Level	0.01	1825	3457	4898	6147	7203	8067	8740	9220	9508	9604	9508	9220	8740	8067	7203	6147	4898	3457	1825
of	0.02	456	864	1225	1537	1801	2017	2185	2305	2377	2401	2377	2305	2185	2017	1801	1537	1225	864	456
Accuracy	0.03	203	384	544	683	800	896	971	1024	1056	1067	1056	1024	971	896	800	683	544	384	203
(d)	0.04	114	216	306	384	450	504	546	576	594	600	594	576	546	504	450	384	306	216	114
	0.05	73	138	196	246	288	323	350	369	380	384	380	369	350	323	288	246	196	138	73
	0.06	51	96	136	171	200	224	243	256	264	267	264	256	243	224	200	171	136	96	51
	0.07	37	71	100	125	147	165	178	188	194	196	194	188	178	165	147	125	100	71	37
	0.08	29	54	77	96	113	126	137	144	149	150	149	144	137	126	113	96	- 77	54	29
	0.09	23	43	60	76	89	100	108	114	117	119	117	114	108	100	89	76	60	43	23
	0.10	18	35	49	61	72	81	87	92	95	96	95	92	87	81	72	61	49	35	18
	0.11	15	29	40	51	60	67	72	76	79	79	79	76	72	67	60	51	40	29	15
	0.12	13	24	34	43	50	56	61	64	66	67	66	64	61	56	50	43	34	24	13
	0.13	11	20	29	- 36	43	48	52	55	56	57	56	55	52	48	43	- 36	29	20	11
	0.14	9	18	25	- 31	37	41	45	47	49	49	49	47	45	41	37	31	25	18	9
	0.15	8	15	22	27	32	36	39	41	42	43	42	41	39	36	32	27	22	15	8
	0.16	7	14	19	24	28	32	34	36	37	- 38	37	36	34	32	28	24	19	14	7
	0.17	6	12	17	21	25	28	30	32	33	33	33	32	30	28	25	21	17	12	6
	0.18	6	11	15	19	22	25	27	28	29	30	29	28	27	25	22	19	15	11	6
	0.19	5	10	14	17	20	22	24	26	26	27	26	26	24	22	20	17	14	10	5
	0.20	5	9	12	15	18	20	22	23	24	24	24	23	22	20	18	15	12	9	5
	0.21	4	8	11	14	16	18	20	21	22	22	22	21	20	18	16	14	11	8	4
	0.22	4	(10	13	15	17	18	19	20	20	20	19	18	17	15	13	10	(4
	0.23	3	(9	12	14	15	17	17	18	18	18	17	17	15	14	12	9	(3
	0.24	3	0	9	11	13	14	15	10	17	17	17	10	15	14	13	11	9	0	3
	0.25	3	0	ð 7	10	12	13	14	10	15	10	15	10	14	13	12	10	8	0	3
	0.20	3	5	7	9	10	12	13	14	14	14	14	14	13	12	10	9	7	5	3
	0.27	3	5	(8	10	10	12	13	13	13	13	13	12	10	10	8	(5	3
	0.28	2	4	0	8	9	10	10	12	12	12	14	12	10	10	9	8	0	4	2
	0.29	2	4	0	7	9	0	10	10	11	11	11	10	10	0	9	7	0	4	2
	0.50	2	4	5	1	ŏ	9	10	10	11	11	11	10	10	9	ð	1	5	4	2

Or use StatCalc (Step 1)

- P = 20% = 0.2 in this example
- D = 5% = 0.05 therefore the true value of the prevalence lies between 15-25%. So worse acceptable result is either 15% or 25%
- Press F4 to calculate.

C:\EPI6\STATCALC.EXE			_ 🗆 🗙
EpiInfo Version 6	Statcalc		November 1993
Population Survey (or Descriptive Study Using	Random (Not Clu	ster) Sampling
Size of popul the sample	lation from which will be selected :	999999	
Expected free under study	(uency of the factor (err toward 50%) :	20.00 ×	
If 50.00 % is the farthest from the or lower)?	e true rate in the populat e rate that you would acce	ion, what is the pt in your sampl	result .e Chigher
Worst	acceptable result :	25.00 ×	
F1-Help	F4-Calc	F6-Open Fi	.le F10-Done

StatCalc (Step 2)

Using 95% confidence level, the sample size required is 246, the same value as manual calculation & the table.

C:\EPI6\STATCALC.EXE					- 🗆 ×
EpiInfo Version 6	Statea	alc		Nove	mber 1993
Population Survey o	r Descriptive Study	Using	g Random (Not	t Cluster)	Sampling
	Population Size	:	999,999		
	Expected Frequency	:	20.00 ×		
	Worst Acceptable	:	2 <mark>5.00</mark> ×		
	Confidence Level	1	Sample Size		
	80 × 90 ×		105 173		
	95 % 99 %		246 < 424		
	99.9 × 99.99 ×		692 968		
Change value of Po	pulation, Frequency,	. or l	Worst Accepta	able to re	calculate.
F1-Help	1	75-Pr :	int F6-Op	en File	F10-Done

© Dr Azmi Mohd Tamil, 2009 Formula for Sample Size of A Prevalence Study

It is the same since all calculations uses the same formula.

C:\EPI6\EPI6.EXE

Chapter 32 - Statistics - p. 393

_ 8 ×

Sample Size Calculations in STATCALC

in which:

The sample size calculations for proportions in a descriptive study or survey use the following method:

Sample size=size of sample randomly selected from the population Population=size of population which the sample is to represent P=true proportion of factor in the population (guess) D=(Maximum) difference between sample mean and population mean Z=area under normal curve corresponding to the desired confidence level.

 Confidence
 Z

 .90
 1.645

 .95
 1.960

 .99
 2.575

 .999
 3.29

The formula is:

Samplesize=n/(1+(n/population))

Esc-Topics

n=Z*Z(P(1-P)/D*D)

0-Quit

SS Calculation for a Known Population

What if the required sample size is larger than the population being studied?

i.e. study on stress among staff at Pusat Kesihatan Rembau. Expected rate of stress is 50% therefore at 5% precision, the required sample size is 384. But the number of staff is only 30!

SS Calculation for a Known Population

Krejcie & Morgan

Krejcie, R.V. & Morgan, D.W. (1970). Determining sample size for research activities. *Educational & Psychological Measurement*, 30, 607-610.

$$S = \frac{X^{2} NP (1 - P)}{d^{2} (N - 1) + X^{2} P (1 - P)}$$

- S = required sample size
 - N = the given population size
 - P = prevalence
 - d = the degree of accuracy
 - $X^2 = 3.841$ for the .95 confidence level

Table - Krejcie, R.V. & Morgan, D.W. (1970).

Rec	uired	Sam	ple	Size
1100	uncu	Juli		OILC.

	Confid	ence = 9	5%		Confid	Confidence = 99%			
Population Size		Margin	of Error			Margin o	of Error		
	5.0%	3.5%	2.5%	1.0%	5.0%	3.5%	2.5%	1.0%	
10	10	10	10	10	10	10	10	10	
20	19	20	20	20	19	20	20	20	
30	28	29	29	30	29	29	30	30	
50	44	47	48	50	47	48	49	50	
75	63	69	72	74	67	71	73	75	
100	80	89	94	99	87	93	96	99	
150	108	126	137	148	122	135	142	149	
200	132	160	177	196	154	174	186	198	
250	152	190	215	244	182	211	229	246	
300	169	217	251	291	207	246	270	295	
400	196	265	318	384	250	309	348	391	
500	217	306	377	475	285	365	421	485	
600	234	340	432	565	315	416	490	579	
700	248	370	481	653	341	462	554	672	
800	260	396	526	739	363	503	615	763	
1,000	278	440	606	906	399	575	727	943	
1,200	291	474	674	1067	427	636	827	1119	
1,500	306	515	759	1297	460	712	959	1376	
2,000	322	563	869	1655	498	808	1141	1785	
2,500	333	597	952	1984	524	879	1288	2173	
3,500	346	641	1068	2565	558	977	1510	2890	
5,000	357	678	1176	3288	586	1066	1734	3842	
7,500	365	710	1275	4211	610	1147	1960	5165	
10,000	370	727	1332	4899	622	1193	2098	6239	
25,000	378	760	1448	6939	646	1285	2399	9972	
50,000	381	772	1491	8056	655	1318	2520	12455	
75,000	382	776	1506	8514	658	1330	2563	13583	
100.000	383	778	1513	8762	659	1336	2585	14227	
250,000	384	782	1527	9248	662	1347	2626	15555	
500,000	384	783	1532	9423	663	1350	2640	16055	
1,000,000	384	783	1534	9512	663	1352	2647	16317	
2,500,000	384	784	1536	9567	663	1353	2651	16478	
10,000,000	384	784	1536	9594	663	1354	2653	16560	
100,000,000	384	784	1537	9603	663	1354	2654	16584	
300,000,000	384	784	1537	9603	663	1354	2654	16586	

Assumption of the table; prevalence = 50%. So need only 28 out of 30 for the study on stress, not 384.

If population > 250,000, sample size equal to Kish's formula.

250,000	384
500,000	384
1,000,000	384
2,500,000	384
10,000,000	384
100,000,000	384
300.000.000	384

† Copyright, The Research Advisors (2006). All rights reserved.

Kish, L (1960) = Krejcie, R.V. & Morgan, D.W. (1970) ?

Kish, L (1960) $n = (Z_{1-\alpha})^2(P(1-P)/D^2)$ S = n/(1+(n/population))

 $(Z_{1-\alpha})^2 = X^2 = 3.841$ Population = N P = P D² = d² = 0.0025 (for 5%)

We usually use only 1st half of the formula!

Krejcie, R.V. & Morgan, D.W. (1970)

$$S = \frac{X^{2} NP (1 - P)}{d^{2} (N - 1) + X^{2} P (1 - P)}$$

So we can use STATCALC to calculate sample size for a known population!

StatCalc

Using 95% confidence level, the sample size required is 28, the same value as in the table.

C:\EPI6\STATCALC.EX	E					- 🗆 ×
EpiInfo Version 6	Statea	alc			Nover	nber 1993
Population Survey	or Descriptive Study	Using	Random	(Not	Cluster)	Sampling
	Population Size	:	30)		
	Expected Frequency	:	50.00) ×		
	Worst Acceptable	:	<u>5</u> 5.00	 ×		
	Confidence Level	S _	ample Si	ze		
	80 × 90 ×		25 27		1	
	99 %		29			
	99.99 %		29			
Change value of F	opulation, Frequency,	. or W	orst Acc	eptal	ble to red	calculate.
F1-Help		75-Pri	nt Fe	-Oper	n File	F10-Done

Calculate Your Own Sample Size

Comparative Cross-Sectional Study – Proving Association Between Risk & Outcome

Comparative Cross-Sectional

What is the outcome being measured?

Is it the pro-

vrisk factor?

 Association between the major risk factor being studied and the outcome?

Cross-Sectional: Risk & Outcome

- If you want to show an association between a risk factor and outcome in a cross-sectional study, then the sample size calculation is different.
- For example you want to prove that Indians (ethnicity = risk) are at higher risk of having diabetes mellitus (outcome) compared to other races in your country, using a crosssectional study.

Example – Indians are of higher risk for DM

- From literature review, identify the rate of disease and proportion of those with the risk factor.
- Proportion of sample from unexposed (Others) population = 85%
- Proportion of sample from exposed (Indians) population = 15%
- P1=true proportion of DM in unexposed (Others) population = 8%
- P2=true proportion of DM in exposed (Indians) population =14%

From Literature Review: Indians & Diabetes M.



Calculate Manually

Calculate using these formulas (Fleiss JL. 1981. pp. 44-45)

$$m' = \frac{\left[c_{\alpha/2}\sqrt{(r+1)\bar{P}\bar{Q}} - c_{1-\beta}\sqrt{rP_1Q_1} + P_2Q_2\right]}{rP_1Q_1}$$

$$r(P_2 - P_1)^2$$

$$1 + \sqrt{1 + \frac{2(r+1)}{m'r|P_2 - P_1|}}^2$$



 $m=n_{1}=size of sample from population 1$ P_1=proportion of disease in population 1 α = "Significance" = 0.05 1-β = Power = 0.8 P = (P_{1}+rP_{2})/(r+1) n_1 = m From table A.2 in Fleiss; If 1- α is 0.95 then c_{α/2} is 1.960 If 1- β is 0.80 then c_{1-beta} is -0.842 n_2 =size of sample from population 2 P_2 =proportion of disease in population 2 β =chance of not detecting a difference = 0.2 r = n2/n1 = ratio of cases to controls $\overline{Q} = 1-\overline{P}$. $n_2 = rm$

2

Calculate Manually

 $P_1 = 0.08$ $P_2 = 0.14$ r = 15/85 $P = (0.08 + (15/85 \ge 0.14))/(1 + 15/85) = 0.089$ Q = 1 - P = 0.911 $\mathbf{m}' = [1.96\sqrt{(15/85+1)x0.089x0.911} - (-0.842)\sqrt{(15/85x0.08x0.92) + (0.14x0.86)}]^2$ $15/85 \ge (0.14 - 0.08)^2$ = 0.8333 = 1311.69760.0006 $\mathbf{m} = \frac{1311.6976}{4} \times \left(\frac{1 + 1 + \frac{(2x(15/85 + 1))}{(1311.6976x15/85x0.06)}}{(1311.6976x15/85x0.06)} \right)$ $|^{2} = 1420.64$ $m = n_1 = 1421$ $n_2 = r \ge m = 15/85 \ge 1421 = 251$ $n_1 + n_2 = 1672$ 56

Or use StatCalc

> Population survey Cohort or cross-sectional Unmatched case-control

n1 = 1422
n2 = 251

Total sample size required is 1673

C:\EPI6\STA	TCALC.EXE				-	
EpiInfo Ve	ersion 6	Statcalc		Novembe	r 1993	3
Unmatche	ed Cohort and Cross-S	ectional Studies	(Exposed and	Nonexpo	sed)	
Probabil: differenc	ity that if the two S ce in the two POPULAT	AMPLES differ thi IONS (Confidence	is reflects a level or 1-α)	true ; :	95.00	×
Probabil: will show	ity that if the two P w a "significant" dif	OPULATIONS diffen ference (Power on	r, the two SAM r 1-β>	IPLES :	80.00	×
Ratio (Nu	umber of Unexposed :	Number of Exposed	i) :	85 :	15	
Expected	frequency of disease	in unexposed gro	ութ	:	8.00	×
Please fi	ll in the closest val	ue to be detected	l for ONE of t	he foll:	owing	:
Risk rat: Odds rat: Percent o	io (RR) or relative r io (OR)closest to 1 disease among exposed	iskclosest to 1 .00 closest to % fo	L.00 or unexposed	ł	1.75 1.87 14.00	×

C:\EPI6\STATCALC.EXE

EpiInfo Version 6

- 🗆 🗙

November 1993

Unmatched Cohort and Cross-Sectional Studies (Exposed and Nonexposed) Sample Sizes for 8.00 % Disease in Unexposed Group

Statcalc

			Disease	Risk	Odds	Sam	ple Size	
Conf.	Power	Unex:Exp	in Exposed	Ratio	Ratio	Unexp.	Exposed	Total
<u>25.00 ×</u>	80.00 ×	85:15	14.00 ×	1.75	1.87	1,422	251	1,673
90.00 ×						1,156	204	1,360
95.00 ×			Change (values f	or	1.422	251	1.673
99.00 ×			inputs a	as desir	ed.	2.023	357	2.380
99.90 %			then pro	ess F4 t	0	2.867	506	3.373
95.00 ×	80.00 ×		recalcu	late.		1.422	251	1.673
0	90.00 ×					1.921	339	2.260
	95.00 ×					2.397	423	2.820
	99.00 ×					3.440	607	4.047
	80.00 ×	1:1				458	458	916
		2:1				666	333	999
		3:1				873	291	1,164
		4:1				1,076	269	1,345
		5:1				1,285	257	1.542
U		6:1				1,488	248	1,736
F1-H	elp			F5-Pri	nt F6-	Open File	F10-De	one

StatCalc = Manual

Sample size calculated manually is similar to the sample size generated by StatCalc.
 Since both methods use Fleiss JL (1981) unequal sample size formula as

shown in the next slide.

The sample size formula for comparison of two groups of equal or unequal size is taken from Fleiss (1), pp. 44-45, formulas 3.19 and 3.20, and table A.2 on p. 259. These formulas do not assume matching of the two populations, and they apply to univariate comparisons of the differences in proportions (rates) in the two populations. m=size of sample from population 1 r*m=size of sample from population 2 P1=true proportion of factor in population 1 P2=true proportion of factor in population 2 alpha=chance of falsely declaring the two proportions to differ ("Significance") beta=chance of not detecting a difference which is present (could be called "Lack of sensitivity") 1-beta=chance of detecting a weal diffemence ("Power") $\frac{\left[c_{\alpha/2}\sqrt{(r+1)\overline{P}\overline{Q}}-c_{1-\beta}\sqrt{rP_1Q_1}+P_2Q_2\right]^2}{rP_1Q_1}$ In the equ These cor to detail correctly $m = \frac{m'}{4} \left[1 + \sqrt{1 + \frac{2(r+1)}{m'r |P_2 - P_1|}} \right]^2 \frac{3.29}{\text{takes considerable attention}}$ The formulas are: m=mprime + (r+1)/(r*:P2-P1:) in which "{" denotes absolute value and: mprime = < < < c [alpha/2]*SQRT1>= < c [1=beta]*SQRT2>>^2>/<r*<< PR2=PR1>^2>> in which "[]" denotes a subscript and: $SQRT1 = ((r+1) \times PB \times QB)^0.5$ SQRT2=((r*PR1*(1-PR1))+(PR2*(1-PR2)))^0.5

Or we can use PS2

Power and	l Sample Size Pi	rogram: Main \	Window					
File Log Help	~	·						
Survival	t-test	Regression 1	Regression 2	Dichotomous	Log			
Output	Output							
What	do you want to know	v? Sam	ple size		•			
<u>Cases</u> <u>chi-sq</u>	ample size for unco uared test	rrected 231						
Design <u>Match</u>	ed or Independent?		Independer	nt	•			
Case (control?		Prospective		•			
How is	the alternative hype	othesis expressed	Z Two propor	tions	•			
Uncor	ected chi-square or	Fisher's exact test	Uncorrected	d chi-square test	•			
Input	<u>a</u> .05	₽ ₀ .08		[Calculate			
<u>ро</u> и	8	₽ ₁ .14 <u>m</u> 5.66666	36661		Liraphs			
Logging is e	nabled.				Exit			

Based on this, the sample size for exposed is 231.

 So unexposed is 85/15*231= 1309
 Total 1540

PS2 ≠ StatCalc

PS2 uses Schlesselman's method for independent case and control groups for studies that will be analysed using an uncorrected chi-square test

- PS2 uses Casagrande et al's method for independent studies that will be analysed using Yates Correction or Fisher's exact test.
- PS2 only uses the generalisation of Casagrande's method proposed by Fleiss for unequal case & control sample size. Even then the answers differ; 1673 vs 1540

Calculate Your Own Sample Size

Case-Control Study

Case-Control

- In a case-control study, you identify the cases and controls. Then you compare the rate of exposure/risk factor between the case and control group.
- For example you want to prove that cataract patients (cases) have a higher rate of diabetes mellitus (risk factor) compared to patients with normal vision (controls).
- From literature review, identify the rate of exposure among the cases (i.e. 50%) and among the controls (i.e. 8%).
- Decide on the ratio; i.e. 1:1

Example – DM higher risk of cataract

- From literature review, identify the rate of risk factor (DM) in cases and controls.
- Proportion of sample from controls (Normal) population = 50%
- Proportion of sample from cases (Cataract) population = 50%
- P1=true proportion of DM in controls (Normal) population = 8%
- P2=true proportion of DM in cases (Cataract) population =50%

From Literature Review: Cataract & Diabetes M.

DM + (50%)



Calculate Manually

Calculate using these formulas (Fleiss JL. 1981. pp. 44-45)

$$m' = \frac{\left[c_{\alpha/2}\sqrt{(r+1)\bar{P}\bar{Q}} - c_{1-\beta}\sqrt{rP_1Q_1} + P_2Q_2\right]}{rP_1Q_1}$$

$$r(P_2 - P_1)^2$$

$$m = \frac{m'}{4} \left[1 + \sqrt{1 + \frac{2(r+1)}{m'r |P_2 - P_1|}} \right]^2$$



m=n₁=size of sample from population 1 P₁=proportion of **exposure** in population 1 α = "Significance" = 0.05 <u>1</u>-β = Power = 0.8 P = (P₁+rP₂)/(r+1) n₁ = m From table A.2 in Fleiss; If 1- α is 0.95 then c_{α/2} is 1.960 If 1- β is 0.80 then c_{1-beta} is -0.842 n₂=size of sample from population 2 P₂=proportion of **exposure** in population 2 β=chance of not detecting a difference = 0.2 r = n2/n1 = ratio of cases to controls Q = 1-P. n₂ = rm

2

Calculate Manually

 $P_1 = 0.08$ $P_2 = 0.5$ r = 1/1

 $P = (0.08 + (1 \ge 0.5))/(1 + 1) = 0.29$ Q = 1 - P = 0.71

 $m' = [1.96\sqrt{(1+1)x0.29x0.71} - (-0.842)\sqrt{(1x0.08x0.92)+(0.5x0.5)}]^2$ = 2.0162 - 17.0001

$$= \underline{3.0163}_{0.1764} = 17.0991$$

$$m = \frac{17.0991}{4} \times \left(\frac{1 + \sqrt{1 + \frac{(2x(1+1))}{(17.0991x1x0.42)}}}{2} \right)^2 = 21.5985$$

 $\mathbf{m} = \mathbf{n}_1 = 22 \qquad \qquad \mathbf{n}_2 = \mathbf{r} \ge \mathbf{m} = 1 \ge 22 \qquad \qquad \mathbf{n}_1 + \mathbf{n}_2 = 44$

Or Use StatCalc

 Tables (2 x 2, 2 x n)

 Sample size & power

 Ch

Population survey Cohort or cross-sectional Unmatched case-control C:\

C:\EPI6\STATCALC.EXE EpiInfo Version 6

So you need a sample size of only 22 cases and 22 controls.

C:\EPI6\STATCALC.EX	L .	- 🗆 🎙
EpiInfo Version 6	Statcalc	November 1993
Unmatched	Case-Control Study (Comparison of	ILL and NOT ILL>
Probability that difference in th	if the two SAMPLES differ this r e two POPULATIONS (Confidence lev	reflects a true vel or 1-α) : 95.00 %
Probability that will show a "sig	if the two POPULATIONS differ, t mificant" difference <power 1-<="" or="" td=""><td>he two SAMPLES β></td></power>	he two SAMPLES β>
#NOT ILL/#ILL <1	means equal sample sizes)	1 : 1
Expected frequen	cy of exposure in NOT ILL group	: 8.00 ×
Please fill in th	e closest value to be detected fo	or ONE of the following:
Odds ratio (OR)- Percent exposure	-closest to 1.00 among ILL groupclosest to % fo	r NOT ILL : 50.00 %

Statcalc	Novem

- 🗆 ×

ber 1993

Unmatched Case-Control Study (Comparison of ILL and NOT ILL) Sample Sizes for 8.00 % Exposure in NOT ILL Group

Conf. 25.00 ×	Power 80.00 %	NOT ILL :ILL 1:1	Exposure in ILL 50.00 %	Odds Ratio 11.50	Sample NOT ILL 22	Size ILL 22	Total 44
90.00 × 95.00 × 99.00 × 99.90 × 95.00 ×	80.00 × 90.00 × 95.00 × 95.00 × 80.00 ×	" " " 4:1 3:1 2:1 1:2 1:3 1:4	Change valu inputs as d then press recalculate	es for esired, F4 to -	18 22 30 43 22 27 32 42 48 39 30 16 15 14	18 22 30 43 22 27 32 42 12 13 15 33 44 55	36 44 60 86 44 54 64 62 45 45 49 59 69
F1-Hel	.p		F5	-Print	F6-Open File	F10	-Done

Or use PS2

Sample size for case is 17 and for control 17. Total 34 StatCalc = 44vs PS2 = 34.....hmmm which would you prefer?

S Power and S	Sample Size Pr	ogram: Ma	ain Wind	low		
ile Log Help						
Survival	t-test	Regression	n 1 Å Re	gression 2	Dichotomous	Log
Output	<u>Stud</u>	ies that are	analyse	d by chi-so	uare or Fisher's o	exact test
What do	you want to knov	<u>12</u>	Sample siz	e		•
<u>Case sar</u> <u>chi-squa</u>	nple size for unco red test	rrected	17			
Design <u>Matched</u>	or Independent?			Independer	nt	•
Case cor	<u>htrol?</u>			Retrospect	ive	•
How is the	ne alternative hypo	thesis expres	sed?	Two propo	rtions	•
Uncorrec	ted chi-square or	Fisher's exac	t test?	Uncorrecte	d chi-square test	•
Input <u>C</u> <u>powe</u>	<u>r</u> .8	$\frac{p_0}{p_1} \begin{bmatrix} .08\\ .08\\ .08\\ .08\\ .08\\ .08\\ .08\\ .08\\$			[*	Graphs
Logging is ena	abled.]	Exit

Calculate Your Own Sample Size

Cohort Study

Cohort

- In a cohort study, you identify those who are currently disease-free. Among them, you identify those with and without the exposure/risk factor. Then this cohort is followed up for a pre-determined amount of time to identify those who developed the disease and those who didn't.
- For example you want to prove that overweight adults have higher risk of diabetes mellitus compared to normal weight adults.

Example – overweight have higher risk of DM

- From literature review, identify the rate of disease among those with & without the risk factor.
- Ratio of unexposed vs exposed; 1:1
- Proportion of sample from no-risk (Normal) population = 50%
- Proportion of sample from at-risk (Overweight) population = 50%
- P1=true proportion of DM in no-risk (Normal) population = 7%
- P2=true proportion of DM in at-risk (Overweight) population =32%


► DM + (32%)



Rifas-Shiman SL et al, 2008. Diabetes and lipid screening among patients in primary care: A cohort study. BMC Health Services Research.

Calculate Manually

Calculate using these formulas (Fleiss JL. 1981. pp. 44-45)

$$m' = \frac{\left[c_{\alpha/2}\sqrt{(r+1)\bar{P}\bar{Q}} - c_{1-\beta}\sqrt{rP_1Q_1} + P_2Q_2\right]}{rP_1Q_1}$$

$$r(P_2 - P_1)^2 - \left[1 + \sqrt{1 + \frac{2(r+1)}{m'r|P_2 - P_1|}}\right]^2$$



m=n₁=size of sample from population 1 P₁=proportion of **disease** in population 1 α = "Significance" = 0.05 <u>1</u>-β = Power = 0.8 P = (P₁+rP₂)/(r+1) n₁ = m From table A.2 in Fleiss; If 1- α is 0.95 then c_{α/2} is 1.960 If 1- β is 0.80 then c_{1-beta} is -0.842

m =

 n_2 =size of sample from population 2 P_2 =proportion of **disease** in population 2 β =chance of not detecting a difference = 0.2 r = n2/n1 = ratio of cases to controls $\overline{Q} = 1-\overline{P}$. $n_2 = rm$

2

$$\begin{array}{c} \textbf{Discrete Manually} \\ P_1 = 0.07 \quad P_2 = 0.32 \quad \mathbf{r} = 1/1 \\ P = (0.07 + (1 \ge 0.32))/(1 + 1) = 0.195 \\ Q = 1 - P = 0.805 \\ \mathbf{m}^* = \underbrace{[1.96\sqrt{(1+1) \ge 0.195 \ge 0.805} - (-0.842)\sqrt{(1 \ge 0.07 \ge 0.93) + (0.32 \ge 0.68)}]^2}_{1 \ge (0.32 - 0.07)^2} \\ = \underbrace{2.3898}_{0.0625} = 38.2369 \\ \mathbf{m} = \underbrace{38.2369}_{4} \ge \underbrace{\left[1 + \sqrt{1 + \underbrace{(2 \ge (1+1)}_{(38.2369 \ge 1 \ge 0.25)}\right]^2} = 45.8882}_{1 \ge 1 \ge 46} \\ \mathbf{m} = \mathbf{n}_1 = 46 \qquad \mathbf{n}_2 = \mathbf{r} \ge \mathbf{m} = 1 \ge 46 = 46 \qquad \mathbf{n}_1 + \mathbf{n}_2 = 92 \end{array}$$

Or Use StatCalc

Tables (2 x 2, 2 x n) Sample size & power

Population survey Cohort or cross-sectional Unmatched case-control

 So you'll need a sample size of 46 each for both groups. Total of 92.

С	:\EPI6\STATC#	ALC.EXE						-	D ,	
Ð	piInfo Vers	ion 6	S	tatcalc		Novem	beı	• 1993	}	
	Unmatched	Cohort and	l Cross-Sect	ional Studie:	s (Exposed a	and Nonex	pos	ed)		
	Probability difference	that if t in the two	he two SAME POPULATION	PLES differ t NS (Confidenc	his reflect: e level or :	s a true 1-α)	: 2	95.00	×	
	Probability will show a	that if t "signific	he two POPU ant" differ	ILATIONS diff rence <power< td=""><td>er, the two or 1-β></td><td>SAMPLES</td><td>: 8</td><td>80.00</td><td>×</td><td></td></power<>	er, the two or 1-β>	SAMPLES	: 8	80.00	×	
	Ratio (Numb	er of Une>	cposed : Nur	nber of Expos	ed) :	1	: 1	L		
	Expected fr	equency of	disease in	n unexposed g	roup		:	7.00	×	
P	lease fill	in the clo	sest value	to be detect	ed for ONE (of the fo	110	wing:		
	Risk ratio Odds ratio	(RR) or re (OR)clos	lative risk est to 1.00	kclosest to	1.00		:	4.57		

C:\EPI6\STATCALC.EXE

EpiInfo Version 6

C:/

-November 1993

: 32.00 ×

- 🗆 ×

Unmatched Cohort and Cross-Sectional Studies (Exposed and Nonexposed) Sample Sizes for 7.00 % Disease in Unexposed Group

Statealc

Percent disease among exposed--closest to % for unexposed

Conf. 25.00 %	Power 80.00 %	Unex:Exp 1:1	Disease in Exposed 32.00 %	Risk Ratio 4.57	Odds Ratio 6.25	Sam Unexp. 46	ple Size Exposed 46	Total 92
90.00 × 95.00 × 99.00 × 99.90 × 95.00 × ""	80.00 × 90.00 × 95.00 × 99.00 × 80.00 ×	" " 4:1 3:1 1:2 1:3 1:4	Change inputs then pr recalcu	values f as desir ess F4 t late.	or ed, o	38 46 65 92 46 58 70 95 100 81 64 36 32 30	38 46 65 92 46 58 70 95 25 27 32 72 97 121	76 92 130 184 92 116 140 125 108 96 108 129 151
F1-H	elp			F5-Pri	.nt F6	-Open File	F10-Da	one

Or use PS2

So the sample size required for each group is 38. Total of 76
 StatCalc = 92 vs PS2 = 76

6	Power and	Sample Size Pr	ogram: Main V	Window					
ile	Log Help								
	Survival	t-test	Regression 1	Regression 2	Dichotomous	Log			
	Output	<u>Stud</u>	ies that are ana	ilysed by chi-so	quare or Fisher's	exact test			
	<u>What do</u>	o you want to know	v? Sam	ple size		-			
	<u>Case sa</u> <u>chi-squa</u>	mple size for unco ared test	rrected 38						
	Design <u>Matched</u>	or Independent?		Independe	nt				
	Case co	ntrol?		Prospectiv	e	-			
	How is t	he alternative hypo	thesis expressed	2 Two propo	Two proportions				
	Uncorre	cted chi-square or	Fisher's exact tes	Uncorrecte	ed chi-square test	•			
	Input	χ	<u>р</u> .07	_		Calculate			
	<u>powe</u>	<u>r</u> .8	P ₁ .32	_	_	Graphs			
			<u>m</u> 1						
L	.ogging is en	abled.				Exit			

Calculate Your Own Sample Size

Clinical Trial Study

Clinical Trial

- Similar approach to cohort if the outcome is categorical.
 But it is easier to refer to available tables.
 For example comparing Fluoxetine against Sertraline for treating depression.
- From literature, 75% of Fluoxetine improved, 70% of Sertraline improved.

Example - treatment of depression

From literature review, identify the rate of improvement in the respective groups.

Ratio of control vs treatment group; 1:1

- Proportion of sample from control (Fluoxetine) population = 50%
- Proportion of sample from treatment (Sertraline) population = 50%
- P1=true proportion of improvement in control (Fluoxetine) population = 75%
- P2=true proportion of improvement in treatment (Sertraline) population =70%

From Literature Review: treatment of depression

Improved (75%)

Fluoxetine
Sample ratio (1:1)
Sertraline
No improvement (25%)
Improved (70%)
No improvement (30%)

Refer to a Table

The fastest way to calculate the sample size is to refer to a table. One such table is published in an article entitled "Clinical Trials in Cancer Research" in Environmental Health Perspectives Vol. 32, pp. 3148, 1979 by Edmund A. Gehan. It is available for download from

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1637 924

Since the cure rate of 75% is not available in the table, we deduct 75% from 100%, giving us 25%. 0.25 is available in the table. The difference of cure rate is 0.05.

For table 3;

- Upper figure: α =0.05, power equals 0.8;
- middle figure: α =0.05, power equals 0.9;
- lower figure: α =0.01, power equals 0.95.

Smaller proportion				larg	er minus	Number of smaller	of patien proporti	ts for va on of su	rious ccess (P	2 – P1) ^a	
(P_1)	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55
0.05	420	130	69	44	36	31	23	20	17	14	13
	570	175	93	59	42	37	31	24	21	18	16
	960	300	155	100	72	54	42	38	33	27	24
0.10	680	195	96	59	41	35	29	23	19	17	13
	910	260	130	79	54	40	36	29	24	20	17
	1550	440	220	135	92	68	52	41	38	32	26
0.15	910	250	120	71	48	39	31	25	20	17	15
-	1220	330	160	95	64	46	40	31	26	22	18
	2060	560	270	160	110	78	59	47	41	35	29
0.20	1090	290	135	80	53	42	33	26	22	18	16
	1460	390	185	105	71	51	43	33	28	23	18
	2470	660	310	180	120	86	64	50	44	36	27
0.25	1250	330	150	88	57	44	35	28	22	18	16
	1680	440	200	115	77	56	45	36	29	23	18
	2840	740	340	200	130	95	68	52	45	36	29
0.30	1380	360	160	93	60	44	36	29	22	18	15
	1840	480	220	125	80	56	46	36	29	23	18
	3120	810	370	210	135	95	69	53	45	36	29
0.35	1470	380	170	96	61	44	36	28	22	17	13
	1970	500	225	130	82	57	46	36	28	22	17
	3340	850	380	215	140	96	69	52	44	35	26
0.40	1530	390	175	97	61	44	35	26	20	17	_
	2050	520	230	130	82	56	45	32	26	20	
	3480	880	390	220	140	95	68	50	41	32	

Table 3. Number of patients needed in an experimental and a control group for a given probability of obtaining a significant result (two-sided test).

Alternative to table

http://www.palmx.org/samplesize/Calc_Samplesize.xls

			m' = [<	$\int_{a/2} \sqrt{(r+1)}$	$\overline{ \overline{P}\overline{Q} } = c_1$ $r(P_2 - c_2)$	$(-\rho \sqrt{rP_1Q})^2$	$P_1 + P_2 Q_2$]5							
	$m = \frac{m^2}{4} \left[1 + \sqrt{1 + \frac{2(r+1)}{m(r)P_2 - P_1}} \right]^2$										idence le	evel 95%	6, Power	80%	
	_	Larger I	Minus Si	maller Pi	roportion	ofSucc	ess (P ₂ -	•P 1)		Ra	atio of ca	ises to c	ontrols :	= 1	
0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.4										0.45	0.50	0.55	0.60	0.65	0.70
Smaller	0.01	249	106	66	47	36	29	23	20	17	15	13	11	10	9
Proportion	0.02	307	120	72	50	38	30	24	20	17	15	13	11	10	9
of	0.03	364	134	77	53	39	31	25	21	18	15	13	11	10	9
Success	0.04	419	147	83	56	41	32	26	21	18	15	13	12	10	9
(P1)	0.05	474	159	88	58	43	33	27	22	18	16	14	12	10	9
	0.06	526	172	93	61	44	34	27	22	19	16	14	12	10	9
	0.07	578	184	98	64	46	35	28	23	19	16	14	12	10	9
	0.08	628	196	103	66	47	36	29	23	19	16	14	12	11	9
	0.09	677	207	108	69	49	37	29	24	20	17	14	12	11	9
	0.10	725	219	112	71	50	38	30	24	20	17	14	12	11	9
	0.11	772	229	117	74	52	39	30	25	20	17	14	12	11	9
	0.12	817	240	121	76	53	40	31	25	20	17	15	12	11	9
	0.13	861	250	125	78	54	40	31	25	21	17	15	13	11	9
	0.14	903	260	130	80	55	41	32	26	21	17	15	13	11	9
	0.15	945	270	133	82	57	42	32	26	21	18	15	13	11	9
	0.16	965	279	137	04	50	43	33	20	21	10	15	13	11	9
	0.17	1024	200	141	88	60	43	33	20	22	18	15	13	11	9
	0.19	10.98	305	148	89	61	44	34	27	22	18	15	13	11	9
	0.20	1133	313	151	91	62	45	34	27	22	18	15	13	11	9
	0.21	1167	321	154	92	63	45	35	27	22	18	15	13	11	9
	0.22	1200	328	157	94	63	46	35	27	22	18	15	13	11	9
	0.23	1231	335	160	95	64	46	35	28	22	18	15	13	11	9
	0.24	1262	342	162	97	65	47	35	28	22	18	15	13	11	9
	0.25	1290	348	165	98	65	47	36	28	22	18	15	13	11	9
	0.26	1318	354	167	99	66	47	36	28	22	18	15	13	11	9
	0.27	1344	360	170	100	67	48	36	28	22	18	15	13	11	9
	0.28	1370	366	172	101	67	48	36	28	22	18	15	12	10	9
	0.29	1393	371	174	102	68	48	36	28	22	18	15	12	10	9
	0.30	1416	376	175	103	68	48	36	28	22	18	15	12	10	9

Calculate Manually

Calculate using these formulas (Fleiss JL. 1981. pp. 44-45)

$$m' = \frac{\left[c_{\alpha/2}\sqrt{(r+1)\bar{P}\bar{Q}} - c_{1-\beta}\sqrt{rP_1Q_1} + P_2Q_2\right]}{rP_1Q_1}$$

$$\frac{r(P_2 - P_1)^2}{4} \left[1 + \sqrt{1 + \frac{2(r+1)}{m'r|P_2 - P_1|}}\right]^2$$



m=n₁=size of sample from population 1 P₁=proportion of **cure** in population 1 α = "Significance" = 0.05 <u>1</u>-β = Power = 0.8 P = (P₁+rP₂)/(r+1) n₁ = m From table A.2 in Fleiss; If 1- α is 0.95 then c_{α/2} is 1.960 If 1- β is 0.80 then c_{1-beta} is -0.842

m = -

n₂=size of sample from population 2 P₂=proportion of **cure** in population 2 β =chance of not detecting a difference = 0.2 <u>r = n2/n1</u> = ratio of treatment grp to controls $\overline{Q} = 1-\overline{P}$.

2

$$n_2 = m_1$$

Calculate Manually

 $P_1 = 0.75$ $P_2 = 0.70$ r = 1/1

 $P = (0.75 + (1 \ge 0.70))/(1 + 1) = 0.725$ Q = 1 - P = 0.275

 $\mathbf{m}^{\prime} = \underline{[1.96\sqrt{(1+1)x0.725x0.275} - (-0.842)\sqrt{(1x0.75x0.25) + (0.70x0.30)}]^2}_{1 \text{ x} (0.75 - 0.70)^2}$

 $= \frac{3.1277}{0.0025} = 1251.0867$

 $m = \frac{1251.09}{4} \times \left(\frac{1 + \sqrt{1 + \frac{(2x(1+1))}{(1251.09x1x0.05)}}}{n_2 = 1290.7769} \right)^2 = 1290.7769$ $m = n_1 = 1291 \qquad n_2 = r \times m = 1 \times 46 = 1291 \qquad n_1 + n_2 = 2582$

Or Use StatCalc

Tables (2 x 2, 2 x n) Sample size & power

Ch

Population survey Cohort or cross-sectional Unmatched case-control

 So you'll need a sample size of 1290 each for both groups. Total of 2580.

C:\EPI6\STATCAL	.C.EXE					- D X					
EpiInfo Versi	on 6	Statcalc		No	vember 19	793					
Unmatched C	Cohort and C	Cross-Sectional Studies	СЕхро	sed and No	nexposed)	,					
Probability difference i	that if the In the two H	two SAMPLES differ th POPULATIONS (Confidence	is ref level	lects a tr or 1-α)	ue : 95.0	10 ×					
Probability will show a	that if the "significan	two POPULATIONS diffent t" difference <power of<="" td=""><td>r, the r 1-β></td><td>two SAMPL</td><td>ES : 80.0</td><td>30 ×</td></power>	r, the r 1-β>	two SAMPL	ES : 80.0	30 ×					
Ratio (Numbe	r of Unexpo	sed : Number of Expose	d> :		1:1						
Expected fre		: 75.0	90 ×								
Please fill in the closest value to be detected for ONE of the following:											
Risk ratio (RR) or relative riskclosest to 1.00 : 0.93 Odds ratio (OR)closest to 1.00 : 0.78 Percent disease among exposedclosest to % for unexposed : 70.00 %											
	Ĩ	-		-							
F1-Help		F4-Calc	F6-	Open File	F10-Da	one					
F1-Help Enilofo Versi	on 6	F4-Calc Statcalc	F6-	Open File No	F10-Do	one 193					
F1-Help EpiInfo Versi Unmatched C Sa	on 6 Cohort and C Imple Sizes	F4-Calc Statcalc Cross-Sectional Studies for 75.00 % Disease in	F6- (Expo Unexp	Open File No sed and No osed Group	F10-Do vember 19 nexposed)	one 993					
F1-Help EpiInfo Versi Unmatched C Sa	on 6 Cohort and C Imple Sizes	F4-Calc Statcalc Cross-Sectional Studies for 75.00 % Disease in Disease Pick O	F6- (Expo Unexp	Open File No sed and No osed Group	F10-Do vember 19 nexposed)	one 793 >					
F1-Help EpiInfo Versi Unmatched C Sa Conf. Power	on 6 Cohort and C Imple Sizes Unex:Exp	F4-Calc Statcalc Cross-Sectional Studies for 75.00 % Disease in Disease Risk O in Exposed Ratio Ra	F6- (Expo Unexp dds tio	Open File No sed and No osed Group Samp Unexp.	F10-Do vember 19 nexposed) le Size Exposed	one 793 Designed					
F1-Help EpiInfo Versi Unmatched C Sa Conf. Power 25.00 × 80.00 ×	on 6 Cohort and Comple Sizes Unex:Exp : 1:1	F4-Calc Statcalc Cross-Sectional Studies for 75.00 % Disease in Disease Risk O in Exposed Ratio Ra 70.00 % 0.93 0	F6- (Expo Unexp dds tio .78	Open File No sed and No osed Group Samp Unexp. 1,290	F10-Do vember 19 nexposed le Size Exposed 1,290	one 293) Total 2,580					
F1-Help EpiInfo Versi Unmatched C Sa Conf. Power 5.00 × 80.00 ×	on 6 Cohort and C Imple Sizes Unex:Exp : 1:1 ''	F4-Calc Statcalc Cross-Sectional Studies for 75.00 % Disease in Disease Risk O in Exposed Ratio Ra 70.00 % 0.93 0	F6- (Expo Unexp dds tio .78	Open File No sed and No osed Group Samp Unexp. 1,290 1,025	F10-Do vember 19 nexposed) le Size Exposed 1,290 1,025	Dine 293 Total 2,580 2,050					
F1-Help EpiInfo Versi Unmatched C Sa Conf. Power 25.00 × 80.00 × 10.00 × "	on 6 Cohort and C Imple Sizes Unex:Exp : 1:1 ''	F4-Calc Statcalc Cross-Sectional Studies for 75.00 % Disease in Disease Risk O in Exposed Ratio Ra 70.00 % 0.93 0 Change values for	F6- (Expo Unexp dds tio .78	Open File No sed and No osed Group Samp Unexp. 1,290 1,025 1,290	F10-Do vember 19 nexposed) le Size Exposed 1,290 1,025 1,290	Dine 293 Total 2,580 2,050 2,80					
F1-Help EpiInfo Versi Unmatched C Sa Conf. Power 5.00 × 80.00 × 9.00 × " 9.00 × "	on 6 Cohort and Comple Sizes Unex:Exp : 1:1 ""	F4-Calc Statcalc Cross-Sectional Studies for 75.00 % Disease in Disease Risk O in Exposed Ratio Ra 70.00 % 0.93 0 Change values for inputs as desired, then press F4 to	F6- (Expo Unexp dds tio .78	Open File No sed and No osed Group Samp Unexp. 1,290 1,025 1,290 1,901 2,762	F10-Do vember 19 nexposed2 le Size Exposed 1,290 1,025 1,290 1,901 2,262	Total 2,050 2,050 3,802 5,524					
F1-Help EpiInfo Versi Unmatched C Sa Conf. Power 15.00 × 80.00 × 19.00 × " 19.00 × " 19.00 × " 19.00 × "	on 6 ohort and 0 mple Sizes Unex:Exp 1:1 ""	F4-Calc Statcalc Cross-Sectional Studies for 75.00 % Disease in Disease Risk O in Exposed Ratio Ra 70.00 % 0.93 0 Change values for inputs as desired, then press F4 to recalculate.	F6- (Expo Unexp dds tio .78	Open File No sed and No osed Group Samp Unexp. 1,290 1,901 2,762 1,290	F10-Do vember 19 nexposed le Size Exposed 1,290 1,025 1,290 1,901 2,762 1,290	Total 2,580 2,580 2,580 3,802 5,524 2,580					
F1-Help EpiInfo Versi Unmatched C Sa Conf. Power 5.00 × 80.00 × 9.00 × " 9.00 × " 9.00 × " 9.00 × " 9.00 × " 9.00 × "	on 6 ohort and 0 mple Sizes Unex:Exp : 1:1 "" ""	F4-Calc Statcalc Cross-Sectional Studies for 75.00 % Disease in Disease Risk Of in Exposed Ratio Ra 70.00 % 0.93 0 Change values for inputs as desired, then press F4 to recalculate.	F6- (Expo Unexp dds tio .78	Open File No sed and No osed Group Unexp. 1,290 1,025 1,290 1,901 2,762 1,290 1,714	F10-Do vember 19 nexposed le Size Exposed 1,290 1,025 1,290 1,901 2,762 1,290 1,214 2,714	Total 2,580 2,050 2,580 3,802 5,524 2,580 3,428 3,428					
F1-Help EpiInfo Versi Unmatched C Sa Conf. Power 5.00 × 80.00 × 9.00 × " 9.00 × " 90.00 × " 90.00 × "	on 6 Cohort and Comple Sizes Unex:Exp 1:1 "" "" "" ""	F4-Calc Statcalc Cross-Sectional Studies for 75.00 % Disease in Disease Risk O in Exposed Ratio Ra 70.00 % 0.93 0 Change values for inputs as desired, then press F4 to recalculate.	F6- (Expo Unexp dds tio .78	Open File No sed and No osed Group Unexp. 1,290 1,025 1,290 1,901 2,762 1,290 1,714 2,110 2,965	F10-Do vember 19 nexposed le Size Exposed 1,290 1,025 1,290 1,901 2,762 1,290 1,714 2,110 2,965	Total 2,580 2,580 2,580 3,802 5,524 2,580 3,428 4,220 5,930					
F1-Help EpiInfo Versi Unmatched C Sa Conf. Power 25.00 × 80.00 × 9.00 × " 9.00 × " 80.00 × " 9.00 × "	on 6 Cohort and Comple Sizes Unex:Exp 1:1 "" "" " " " " " " " " " " "	F4-Calc Statcalc Cross-Sectional Studies for 75.00 % Disease in Disease Risk O in Exposed Ratio Ra 70.00 % 0.93 Ø Change values for inputs as desired, then press F4 to recalculate.	F6- (Expo Unexp dds tio .78	Open File No sed and No osed Group Unexp. 1,290 1,025 1,290 1,901 2,762 1,290 1,714 2,10 2,965 3,180	F10-Do vember 19 nexposed le Size Exposed 1,290 1,025 1,290 1,901 2,762 1,290 1,714 2,110 2,965 795	Total 2,580 2,580 2,580 3,802 5,524 2,580 3,802 5,524 2,580 3,802 5,524 2,580 3,802 3,975					
F1-Help EpiInfo Versi Unmatched C Sa Conf. Power 25.00 × 80.00 × 9.00 × " 9.00 × "	on 6 ohort and 0 mple Sizes Unex:Exp 1:1 " " " " " " " " " " " " " " " " " "	F4-Calc Statcalc Cross-Sectional Studies for 75.00 % Disease in Disease Risk O in Exposed Ratio Ra 70.00 % 0.93 0 Change values for inputs as desired, then press F4 to recalculate.	F6- (Expo Unexp dds tio .78	Open File No sed and No osed Group Samp Unexp. 1,290 1,025 1,290 1,290 1,2762 1,290 1,2762 1,290 1,714 2,110 2,965 3,180 2,550	F10-Do vember 19 nexposed le Size Exposed 1,290 1,901 2,762 1,290 1,714 2,110 2,965 795 850 966	Total 2,050 2,050 2,580 3,802 5,580 3,802 5,580 3,428 4,220 5,930 3,975 3,400					
F1-Help EpiInfo Versi Unmatched C Sa Conf. Power 25.00 × 80.00 × 19.00 × " 19.00 × " 10.00 × " 1	on 6 ohort and 0 unple Sizes Unex:Exp 1:1 " " " " " " " " " " " " " " " " " "	F4-Calc Statcalc Cross-Sectional Studies for 75.00 % Disease in Disease Risk O in Exposed Ratio Ra 70.00 % 0.93 0 Change values for inputs as desired, then press F4 to recalculate.	F6- (Expo Unexp dds tio .78	Open File No sed and No osed Group Samp Unexp. 1,290 1,901 2,762 1,290 1,901 2,762 1,290 1,714 2,110 2,965 3,180 2,550 1,920 974	F10-Do vember 19 nexposed 1,290 1,025 1,290 1,901 2,762 1,290 1,714 2,110 2,965 795 850 960 1,949	Total 2,580 2,580 2,580 3,802 5,524 2,580 3,428 4,220 5,930 3,975 3,400 2,880 2,880 2,923					
F1-Help EpiInfo Versi Unmatched C Sa Conf. Power 25.00 × 80.00 × 90.00 × 90.0	on 6 Cohort and Comple Sizes Unex:Exp 1:1 " " " " " " " " " " " " " " " " " "	F4-Calc Statcalc Cross-Sectional Studies for 75.00 % Disease in Disease Risk Of in Exposed Ratio Ra 70.00 % 0.93 0 Change values for inputs as desired, then press F4 to recalculate.	F6- (Expo Unexp dds tio .78	Open File No sed and No osed Group Samp Unexp. 1,290 1,025 1,290 1,901 2,762 1,290 1,901 2,762 1,290 1,714 2,110 2,965 3,180 2,550 1,920 974 869	F10-Do vember 19 nexposed 1,290 1,025 1,290 1,901 2,762 1,290 1,714 2,100 2,965 795 850 960 1,949 2,607	Total 2,580 2,050 2,580 3,802 5,524 2,580 3,428 4,220 3,975 3,400 2,880 2,923 3,476					
F1-Help EpiInfo Versi Unmatched C Sa Conf. Power 5.00 × 80.00 × 9.00 × " 9.00 × " 0.00 ×	on 6 Cohort and Comple Sizes Unex:Exp 1:1 " " " " " " " " " " " " " " " " " "	F4-Calc Statcalc Cross-Sectional Studies for 75.00 % Disease in Disease Risk Or in Exposed Ratio Ra 70.00 % 0.93 0 Change values for inputs as desired, then press F4 to recalculate.	F6- (Expo Unexp dds tio .78	Open File No sed and No osed Group Samp Unexp. 1,290 1,025 1,290 1,901 2,762 1,290 1,714 2,165 3,180 2,550 1,920 974 869 816	F10-Do vember 19 nexposed 1,290 1,025 1,290 1,901 2,762 1,290 1,91 2,762 1,290 1,91 2,762 1,290 1,91 2,762 1,290 1,91 2,762 1,290 1,949 2,607 3,265	Total 2,580 2,580 2,580 2,580 3,802 5,524 2,580 3,428 4,220 3,428 4,220 3,975 3,400 2,880 2,923 3,476 4,081					

Or use PS2

So the sample size required for each group is 1251. Total of 2502 StatCalc = 2580 vs PS2 = 2502

Ś	Power and	Sample Size Pr	ogram: N	Aain Wi	ndow						
ile	Log Help										
	Survival	t-test	Regressi	on 1	Regression 2	Dichotomous	Log				
	Output	<u>Stud</u>	<u>ies that ar</u>	e analy	sed by chi-sc	uare or Fisher's	exact test				
	<u>What do</u>) γou want to knov	<u>v?</u>	Sample	e size		•				
	<u>Case sa</u> <u>chi-squa</u>	mple size for unco red test	rrected	1251							
	Design										
	Matched	or Independent?			Independer	nt	<u> </u>				
	Case cor	ntrol?			Prospective	Prospective					
	How is t	he alternative hypo	othesis expre	essed?	Two propo	Two proportions Uncorrected chi-square test					
	Uncorrec	cted chi-square or	Fisher's exa	act test?	Uncorrecte						
	Input	<u>v</u> .05	p_	75			Calculate				
	<u>powe</u>	<u>r</u> .8	p_1	70	-		Graphs				
			<u> </u>								
L	.ogging is en	abled.					Exit				

Table vs StatCalc vs PS2

From table; 1250 from each group = 2500.
 From PS2; 1251 from each group = 2502
 From StatCalc; 1290 from each group = 2580.

- From manual calculation; 1291 from each group = 2582.
- So the sample size from the table is very similar to PS2's results.

What if the outcome is continuous data?

Jones SR, Carley S & Harrison M. An introduction to power and sample size estimation. Emergency Medical Journal 2003;20;453-458. 2003

Continuous data (two independent groups)

We need to specify the following;

- Standard deviation of the variable (s.d)
- Clinically relevant difference (δ)
- The significant level $(\alpha) 0.05$
- The power $(1 \beta) 80\%$

Continuous data (two independent groups)

The standardised difference is calculated as;

S.d

Example

If difference between means = 10 mmHg
 & pop. standard deviation = 20 mm Hg
 Then standardised difference;

10 mm Hg/20 mm Hg = 0.5

Continuous data (two independent groups)

We draw a straight line from the value for the standardized difference to the value of 0.80 on the scale for power.
 Read off the value for N on the line corresponding to α = 0.05, which gives a total sample size of eg. 128, so we required 64 samples for each group.



Or refer to a table

 Sdiff = 0.5, sample size = 64.
 So 2 groups = 128. **Table 3**How power changes with standardiseddifference

	Power lev	Power level (pβ)			
Sdiff	0.99	0.95	0.90	0.80	
0.10	3676	2600	2103	1571	
0.20	920	651	527	394	
0.30	410	290	235	176	
0.40	231	164	133	100	
0.50	148	105	86	64	
0.60	104	74	60	45	
0.70	76	54	44	33	
0.80	59	42	34	26	
0.90	47	34	27	21	
1.00	38	27	22	17	
1.10	32	23	19	14	
1.20	27	20	16	12	
1.30	23	17	14	11	
1.40	20	15	12	9	
1.50	18	13	11	8	

Sdiff, standardised difference.

Alternative to table

http://www.palmx.org/samplesize/Calc_Samplesize.xls

Sdiff =	ð	δ = Clinically relevant difference							
	s.d	s.d = Stand	dard deviation	on of the va	riable				
Confide	nce level	99	9%	95%					
Powe	r Level	0.90	0.80	0.90	0.80				
	0.10	2977	2337	2103	1571				
	0.20	745	585	527	394				
	0.30	332	261	235	175				
	0.40	187	147	132	99				
	0.50	120	94	85	64				
	0.60	84	66	59	45				
	0.70	62	49	44	33				
Sdiff	0.80	48	38	34	26				
	0.90	38	30	27	20				
	1.00	31	24	22	17				
	1.10	26	20	18	14				
	1.20	22	17	16	12				
	1.30	19	15	13	10				
	1.40	16	13	12	9				
	1.50	14	11	10	8				

©drtamil@medic.ukm.my

Azmi M.T. 2008. Calculate Your Own Sample Size . Kuala Lumpur: Penerbit SPF

Or you can use PS2

 We still end up with the same answer.

è\$	Powe	er and	Sample Size P	rogram: Main	Window				×
File	Log	Help							
Ĺ	Surv	vival	t-test	Regression 1	Regressi	on 2 1	Dichotomous	Log	
	Outp	out			Stuc	lies that	are analyse	d by t-tests	
		What do	o you want to know	w? Sar	nple size			•	
		Sample	<u>Size</u>	64			1		
	Desi	gn							
		<u>Paired o</u>	r independent?			Indepen	dent	•	
	Inpu	t							
		2	<u>x</u> .05	<u>δ</u> [σ [10			Calculate	
		powe	<u>r</u> .8	<u>m</u>	1			Graphs	
L	.oggin	ıg is en	abled.					Exit	

Manual Calculation

$$n = 1 + 2C\left(\frac{s}{d}\right)$$

(Snedecor and Cochran 1989)

s = standard deviation,
d = the difference to be detected, and
C = constant (refer to table below); if α=0.05 & 1-β=0.8, then C = 7.85.

1-β 0.8 7.85 11.68

0.9 10.51 14.88

Manual Calculation

d = 10 mmHg
 s = 20 mm Hg

```
n = 1 + 2 \times 7.85 (20/10)^2
= 63.8 = 64
```

This is similar to the table and PS2!

Calculate Your Own Sample Size

Diagnostic Study

Jones SR, Carley S & Harrison M. An introduction to power and sample size estimation. Emergency Medical Journal 2003;20;453-458. 2003

Diagnostic Study

Based upon literature review, identify the sensitivity & specificity of the diagnostic test being studied.
 Calculate the sample size based on the sensitivity using the following formula;

$$TP + FN = z^{2} \times \frac{(SN(1 - SN))}{W^{2}}$$
$$N(sN) = \frac{TP + FN}{P}$$

Terms

- TP = True Positive FN = False Negative SN = Sensitivity z = Confidence Interval normal distribution value i.e. for 95%, z = 1.96P = prevalence of disease in the test population
- W = accuracy = 0.05

Example (SN=95%)

SN = 95%
z = 1.96
P = 30%
W = 0.05

Example (SN=95%)

$$TP + FN = z^{2} \times \frac{(SN(1 - SN))}{W^{2}}$$
$$= 1.96^{2} \times \frac{(0.95(1 - 0.95))}{0.05^{2}}$$
$$= 3.842 \times \frac{0.0475}{0.0025}$$
$$N(sN) = \frac{TP + FN}{P} = \frac{72.998}{0.2}$$

Ρ

0.3

Diagnostic Study

Calculate again the sample size based on the specificity using the following formula;

$$FP + TN = z^{2} \times \frac{(SP(1 - SP))}{W^{2}}$$
$$N(sp) = \frac{FP + TN}{(1 - P)}$$

Example (SP=80%)

SP = 80%
z = 1.96
P = 30%
W = 0.05

Example (SP=80%)

$$FP + TN = z^{2} \times \frac{(SP(1 - SP))}{W^{2}}$$
$$= 1.96^{2} \times \frac{(0.80(1 - 0.80))}{0.05^{2}}$$

$$= 3.842 \times \frac{0.16}{0.0025}$$

$$N(sp) = \frac{FP + TN}{(1 - P)} = \frac{245.888}{(1 - 0.3)}$$
Diagnostic Study

N for Sensitivity 95% = 243
N for Specificity 80% = 351
If interested in both sensitivity and specificity, then take the higher number (e.g. 351).

Conclusion

- You can calculate your own sample size.
- Tools are available and most of them are free.
 - Decide what is your study design and choose the appropriate method to calculate the sample size.

If despite following these notes fastidiously, your proposal is still rejected by the committee due to sample size, kindly SEE THEM, not us.

References (incl. for StatCalc)

- Fleiss JL. Statistical methods for rates and proportions. New York: John Wiley and Sons, 1981.
- Gehan EA. Clinical Trials in Cancer Research. Environmental Health Perspectives Vol. 32, pp. 3148, 1979.
- Jones SR, Carley S & Harrison M. An introduction to power and sample size estimation. Emergency Medical Journal 2003;20;453-458. 2003

Kish L. Survey sampling. John Wiley & Sons, N.Y., 1965.

- Krejcie, R.V. & Morgan, D.W. (1970). Determining sample size for research activities. *Educational & Psychological Measurement*, 30, 607-610.
- Snedecor GW, Cochran WG. 1989. Statistical Methods. 8th Ed. Ames: Iowa State Press.

© Dr Azmi Mohd Tamil, 2009

References (PS2)

- Dupont WD, Plummer WD, Jr: Power and Sample Size Calculations: A Review and Computer Program. Controlled Clinical Trials 11:116-128, 1990
- Dupont WD, Plummer WD, Jr: Power and Sample Size Calculations for Studies Involving Linear Regression. Controlled Clinical Trials 19:589-601, 1998
- Schoenfeld DA, Richter JR: Nomograms for calculating the number of patients needed for a clinical trial with survival as an endpoint. Biometrics 38:163-170, 1982
- Pearson ES, Hartley HO: Biometrika Tables for Statisticians Vol. I 3rd Ed. Cambridge: Cambridge University Press, 1970
- Schlesselman JJ: Case-Control Studies: Design, Conduct, Analysis. New York: Oxford University Press, 1982
- Casagrande JT, Pike MC, Smith PG: An improved approximate formula for calculating sample sizes for comparing two binomial distributions. Biometrics 34:483-486, 1978
- Dupont WD: Power calculations for matched case-control studies. Biometrics 44:1157-1168, 1988
- Fleiss JL. Statistical methods for rates and proportions. New York: John Wiley and Sons, 1981.

