临床诊断实验样本量估算

作者：王晓晓，赵一鸣 来源：临床流行病学和循证医学

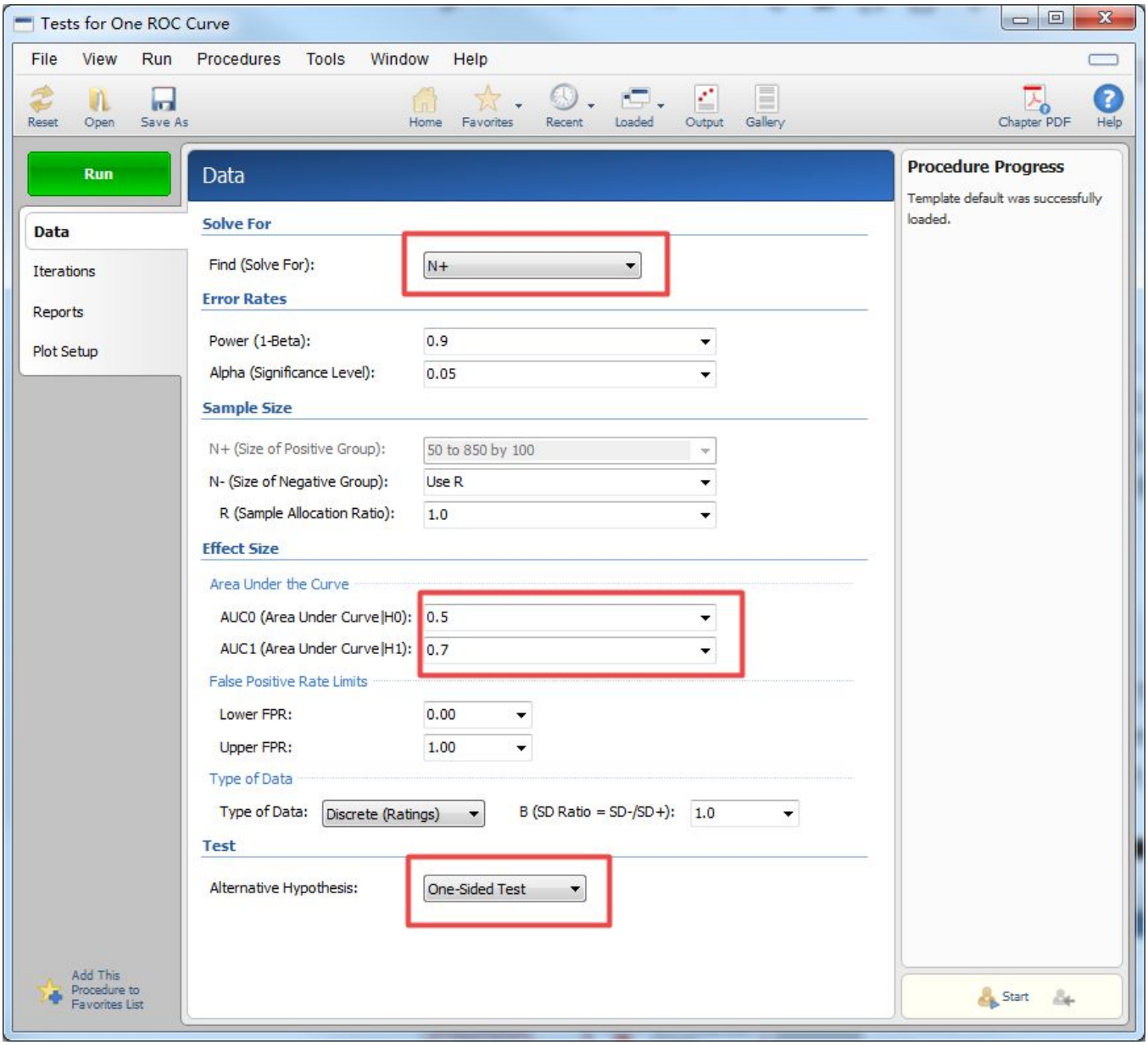
本文原地址：https://www.iikx.com/news/statistics/5917.html

临床诊断实验样本量估算。有这样两份标书，凝练之后，主要目标如下：(1)研究指标A，能否有效区分患者和健康对照;(2)研究工具A，能否有效预测并发症发生与否。

可以看到，申请者主要关注某个指标的诊断/预测价值，那怎么估算样本量呢?我们知道，评价诊断或预测价值时，可以选择如下指标：ROC曲线下面积(AUC)或灵敏度、特异度。

第一种情况，以AUC为主要指标估算样本量

研究假说是：待评价指标具备较好的诊断价值，AUC大于0.5。可以借助PASS11软件中“ DiagnosticTest(ROC) ”---“ Testfor One ROC Curve ”模块进行样本量估算。


Tests for One ROC Curve
File View Run Procedures Tools Window Help
Reset Open Save As Home Favorites Recent Loaded Output Gallery Chapter PDF Help
Run
Data
Iterations
Reports
Plot Setup
Data
Solve For
Find (Solve For): N+
Error Rates
Power (1-Beta): 0.9
Alpha (Significance Level): 0.05
Sample Size
N+ (Size of Positive Group): 50 to 850 by 100
N- (Size of Negative Group): Use R
R (Sample Allocation Ratio): 1.0
Effect Size
Area Under the Curve
AUC0 (Area Under Curve|H0): 0.5
AUC1 (Area Under Curve|H1): 0.7
False Positive Rate Limits
Lower FPR: 0.00
Upper FPR: 1.00
Type of Data
Type of Data: Discrete (Ratings)
B (SD Ratio = SD-/SD+): 1.0
Test
Alternative Hypothesis: One-Sided Test
Procedure Progress
Template default was successfully loaded.
Add This Procedure to Favorites List
Start

One ROC Curve Power Analysis

Numeric Results for Testing AUC0 = AUC1 with Discrete (Rating) Data

Test Type = One-Sided. FPR1 = 0.000. FPR2 = 1.000. B = 1.000. Allocation Ratio = 1.000.

Power	N+	N-	AUC0'	AUC1'	Diff'	AUC0	AUC1	Diff	Alpha	Beta
0.90152	33	33	0.5000	0.7000	0.2000	0.5000	0.7000	0.2000	0.05000	0.09848

References

Hanley, J. A. and McNeil, B. J. 1983. 'A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases.' Radiology, 148, 839-843. September, 1983.

Obuchowski, N. and McClish, D. 1997. 'Sample Size Determination for Diagnostic Accuracy Studies Involving Binormal ROC Curve Indices.' Statistics in Medicine, 16, pages 1529-1542.

Report Definitions

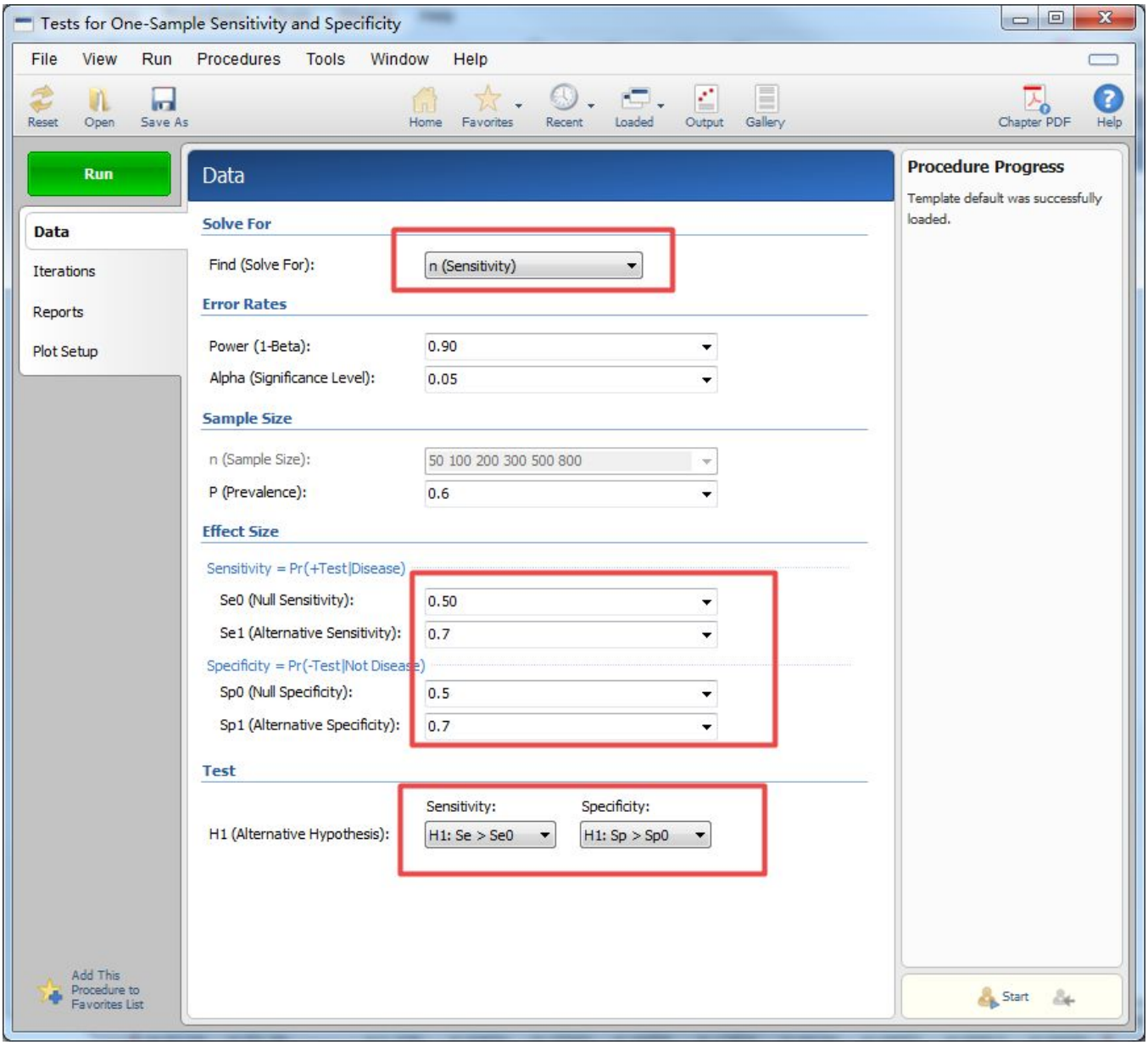
Power is the probability of rejecting a false null hypothesis.
N+ is the sample size from the positive (diseased) population.
N- is the sample size from the negative (non-diseased) population.
Alloc Ratio is the Sample Allocation Ratio (R = N- / N+).
AUC0' is the adjusted area under the ROC curve under the null hypothesis.
AUC1' is the adjusted area under the ROC curve under the alternative hypothesis.
Diff' is AUC1 - 'AUC0. This is the adjusted difference to be detected.
AUC0 is the actual area under the ROC curve under the null hypothesis.
AUC1 is the actual area under the ROC curve under the alternative hypothesis.
Diff is AUC1 - AUC0. This is the difference to be detected.
Alpha is the probability of rejecting a true null hypothesis.
Beta is the probability of accepting a false null hypothesis.
FPR1, FPR2 are the lower and upper bounds on the false positive rates.
B is the ratio of the standard deviations of the negative and positive groups.

Summary Statements

A sample of 33 from the positive group and 33 from the negative group achieve 90% power to detect a difference of 0.2000 between the area under the ROC curve (AUC) under the null hypothesis of 0.5000 and an AUC under the alternative hypothesis of 0.7000 using a one-sided z-test at a significance level of 0.05000. The data are discrete (rating scale) responses. The AUC is computed between false positive rates of 0.000 and 1.000. The ratio of the standard deviation of the responses in the negative group to the standard deviation of the responses in the positive group is 1.000.

标书或者文章中可以这么写：本研究假说为指标A能有效区分患者和健康对照，研究假说为：指标A的ROC曲线下面积大于0.5。前期预实验(或查文献)获知指标A的ROC曲线下面积为0.7，在α=0.05(单侧)，β=0.1，组间比例1:1，采用PASS11估算样本量。结果发现，至少需要纳入患者33人，对照33人。考虑10%的失访率，研究纳入患者37人，对照37人。

第二种情况，采用灵敏度和特异度评价诊断或预测价值。可以借助PASS11软件中“DiagnosticTest(ROC)”---“Testfor One Sample Sensitivity and Specificity”模块进行样本量估算。


Tests for One-Sample Sensitivity and Specificity
File View Run Procedures Tools Window Help
Reset Open Save As Home Favorites Recent Loaded Output Gallery Chapter PDF Help
Run
Data
Iterations
Reports
Plot Setup
Data
Solve For
Find (Solve For):
n (Sensitivity)
Error Rates
Power (1-Beta):
0.90
Alpha (Significance Level):
0.05
Sample Size
n (Sample Size):
50 100 200 300 500 800
P (Prevalence):
0.6
Effect Size
Sensitivity = Pr(+Test|Disease)
Se0 (Null Sensitivity):
0.50
Se1 (Alternative Sensitivity):
0.7
Specificity = Pr(-Test|Not Disease)
Sp0 (Null Specificity):
0.5
Sp1 (Alternative Specificity):
0.7
Test
Sensitivity:
Specificity:
H1 (Alternative Hypothesis):
H1: Se > Se0
H1: Sp > Sp0
Procedure Progress
Template default was successfully loaded.
Add This Procedure to Favorites List
Start

One-Sample Sensitivity and Specificity Power Analysis

Numeric Results for testing H0: Se = Se0 vs. H1: Se > Se0 and H0: Sp = Sp0 vs. H1: Sp > Sp0
Test Statistic: Binomial Test

----- Power -----			--- Sensitivity ---		--- Specificity ---		---------- Alpha ----------			
Sens.	Spec.	Sample Size N1 and N	H0 Se0	H1 Se1	H0 Sp0	H1 Sp1	Target	Sens. Actual	Spec. Actual	Prevalence P
0.9138	0.9138	53 106	0.5000	0.7000	0.5000	0.7000	0.0500	0.0492	0.0492	0.5000

References

Obuchowski, N.A., Zhou, X.H. 2002. 'Prospective studies of diagnostic test accuracy when disease prevalence is low,' Biostatistics, Volume 3, No. 4, pages 477-492.

Li, J., Fine, J. 2004. 'On sample size for sensitivity and specificity in prospective diagnostic accuracy studies,' Statistics in Medicine, Volume 23, pages 2537-2550.

Machin, D., Campbell, M.J., Tan, S.B., Tan, S.H. 2008. Sample Size Tables for Clinical Studies, Third Edition. Wiley-Blackwell, Chichester, United Kingdom.

Zhou, X.H., Obuchowski, N.A., McClish, D.K. 2002. Statistical Methods in Diagnostic Medicine. Wiley-Interscience, New York.

Report Definitions

Sens. Power is the power of the sensitivity test. It is based on the N1 observations.
Spec. Power is the power of the specificity test. It is based on the N-N1 observations.
N is the total sample size of the study. It is equal to N1 + N2.
Se0 is the sensitivity under H0. The sensitivity is the proportion of diseased subjects that yield a positive test result.
Se1 is the sensitivity under H1. The sensitivity is the proportion of diseased subjects that yield a positive test result.
Sp0 is the specificity under H0. The specificity is the proportion of non-diseased subjects that yield a negative test result.
Sp1 is the specificity under H1. The specificity is the proportion of non-diseased subjects that yield a negative test result.
Target Alpha is the alpha (probability of rejecting H0 when H0 is true) that was desired.
Actual Sens. Alpha is the alpha that was actually achieved by the sensitivity test, calculated from the binomial distribution.
Actual Spec. Alpha is the alpha that was actually achieved by the specificity test, calculated from the binomial distribution.
P is proportion of the population that actually has the condition (disease) of interest, called the prevalence.

Summary Statements

A total sample size of 106 (which includes 53 subjects with the disease) achieves 91% power to detect a change in sensitivity from 0.5 to 0.7 using a one-sided binomial test and 91% power to detect a change in specificity from 0.5 to 0.7 using a one-sided binomial test. The target significance level is 0.05. The actual significance level achieved by the sensitivity test is 0.0492 and achieved by the specificity test is 0.0492. The prevalence of the disease is 0.5.

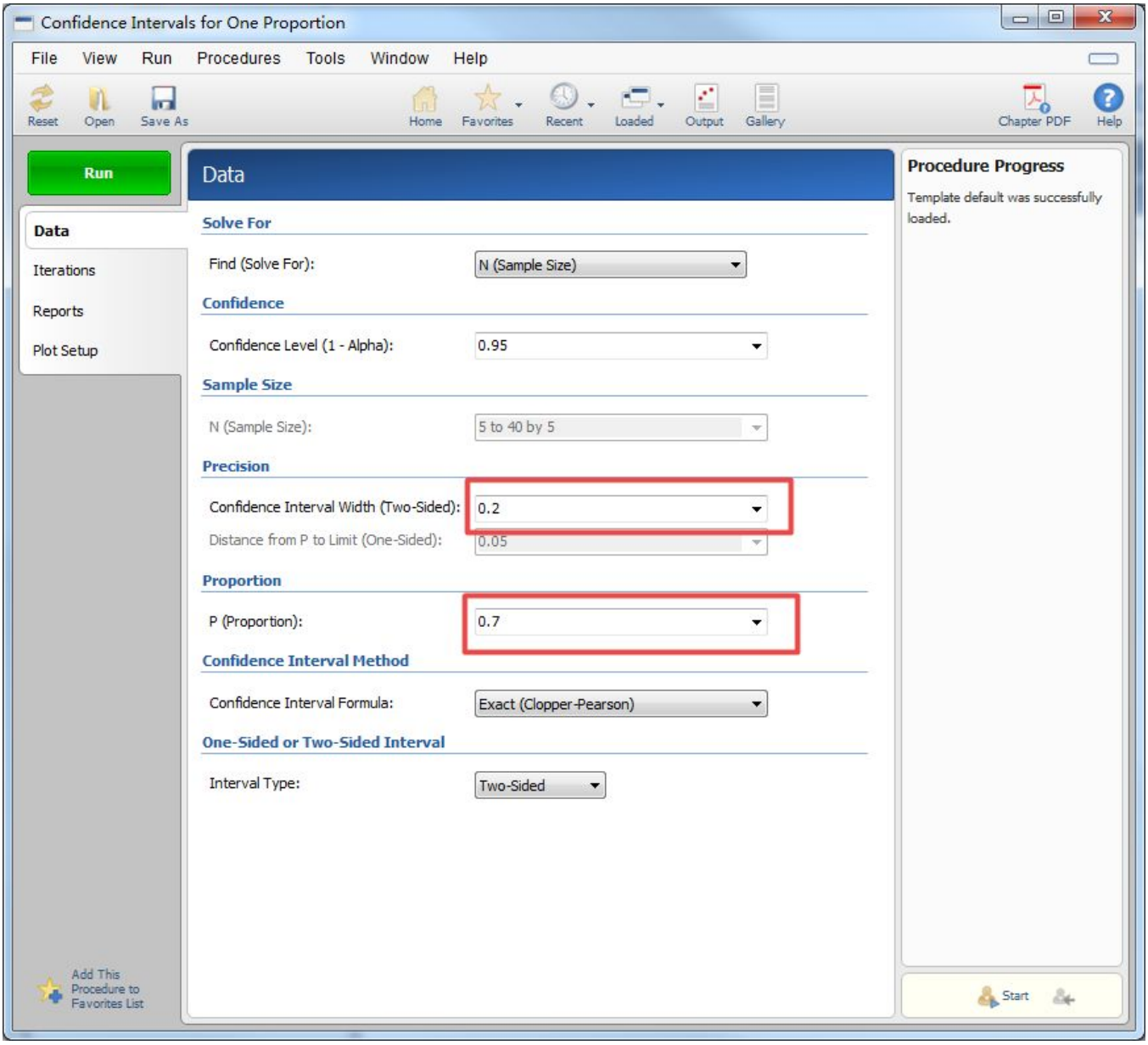
标书或者文章中可以这么写：本研究假说为指标A能有效区分患者和健康对照，研究假说为：指标A的灵敏度和特异度均大于0.5。前期预实验(或查文献)获知指标A的灵敏度为0.7，特异度为0.7，在α=0.05(单侧)，β=0.1，组间比例1:1(患病率为0.5)，采用PASS11估算样本量。结果发现，至少需要纳入106人，其中患者53人，对照53人。考虑10%的失访率，研究纳入患者59人，对照59人。

(患病率P指的是研究对象中患者的比例，如果患者和健康对照的比例为1:1，则P应为0.5。如果患者和健康对照的比例为2:1，则P约为0.67。)

第三种情况，通过估算灵敏度的置信区间获知患者样本量

，通过估算特异度的置信区间估算对照组样本量。可以借助PASS11软件中“ ConfidenceIntervals ”

---“ Proportions ”---“ Confidence Intervals for One Proportion ”模块进行样本量估算。



Confidence Intervals for One Proportion - New

Numeric Results for Two-Sided Confidence Intervals for One Proportion

Confidence Interval Formula: Exact (Clopper-Pearson)

Confidence Level	Sample Size (N)	Target Width	Actual Width	Proportion (P)	Lower Limit	Upper Limit	Width if P = 0.5
0.950	89	0.200	0.199	0.700	0.594	0.793	0.216

References

Fleiss, J. L., Levin, B., Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.

Newcombe, R. G. 1998. 'Two-Sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods.' Statistics in Medicine, 17, pp. 857-872.

Report Definitions

Confidence level is the proportion of confidence intervals (constructed with this same confidence level, sample size, etc.) that would contain the population proportion.
N is the size of the sample drawn from the population.
Width is the distance from the lower limit to the upper limit.
Target Width is the value of the width that is entered into the procedure.
Actual Width is the value of the width that is obtained from the procedure.
Proportion (P) is the assumed sample proportion.
Lower Limit is the lower limit of the confidence interval.
Upper Limit is the upper limit of the confidence interval.
Width if P = 0.5 is the maximum width for a confidence interval with sample size N.

Summary Statements

A sample size of 89 produces a two-sided 95% confidence interval with a width equal to 0.199 when the sample proportion is 0.700.

标书或者文章中可以这么写：本研究假说为指标A对于术后并发症的发生具有一定的预测价值。查文献获知，A的灵敏度=0.7，特异度=0.8。通过估算灵敏度和特异度的可信区间来估算样本量，在α=0.05，置信区间的宽度为0.2，采用PASS11计算样本量分别为89人和70人。要求至少89人发生并发症，至少70人未发生并发症。预实验发现并发症发生率为60%，则需纳入175人(89/0.6=149,70/0.4=175)，才能保证并发症发生人数不低于89人，未发生并发症的人数不低于70人。考虑10%的失访率，纳入患者193人。