

Comparison of Three Methods in Calculating Confidence Intervals for Rate Differences in Non-inferior Clinical Trials

Yaohua Huang, PPD, Beijing, China

ABSTRACT

Non-inferior clinical trials are widely used in pharmaceutical and medical device products development. While the primary endpoints are categorical variables, confidence intervals for rate differences are usually calculated. These confidence intervals are compared with pre-specified non-inferior margin to judge whether non-inferiority assumption can be reached. Approximately Normal, Exact Probability, and Newcombe-Wilson Score methods are introduced in this paper to calculate the confidence intervals for rate differences. In addition, power analysis of these methods will be conducted with Monte Carlo simulation. If sample size is planned using Approximately Normal, all 3 methods can provide enough power if at least one group's rate is neither 0% or 100%. If both groups' rates are 0% or 100% only Newcombe-Wilson Score method can provide the confidence interval but the power is not sufficient.

Key words: Confidence Intervals, Newcombe-Wilson Score, Power

INTRODUCTION

In some drug and medical device clinical trials, some products (such as Orthopaedics) always have the success, efficient rate of 100% because of advanced technique. If this kind of studies are randomized controlled trials, sponsor often conservatively estimate product's success rate, like 98%, 99% in trial design. Sample size is calculated with the estimate rate considering non-inferior margin recognized by clinician. However, if there is no failure case in the final results, namely both of experiment and control groups have the rate of 100%, confidence intervals can't be estimated using traditional CMH (Cochran-Mantel-Haenszel) chisquare method due to unable to estimate rate intervals' standard error. Therefore it's urgent to have a new evaluation method to calculate the rate difference when both of rates are 100%.

METHODS

Non-inferiority clinical trial, it's a kind of trial evaluating the efficacy and safety of product by comparison with positive controlled group. This study trial design is widely used in drug and medical device clinical trials. For the category variables, the comparisons can be transformed into the difference between incidence difference of treatment and controlled group with pre-specified margin. With the assumption that the primary endpoint is success rate, such as success rate, hypothesis $H_0: P_T - P_C \geq \Delta$, $H_1: P_T - P_C < \Delta$, P_T is success rate of treatment group, P_C is success rate of controlled group, Δ is pre-specified non-inferior margin, it's decided by clinician and statistician.

Normal approximation

In the stage of trial design, normal approximation is usually used to calculate sample size, the formula of sample size is

$$n = \frac{\left[z_{1-\alpha/2} \sqrt{2\bar{p}(1-\bar{p})} + z_{1-\beta} \sqrt{p_1(1-p_1) + p_2(1-p_2)} \right]^2}{(\Delta - (p_1 - p_2))^2}$$

$\bar{p}$ is average incidence rate of both groups, Δ is non-inferiority margin.

In the stage of evaluation, normal approximation method is always used to evaluate the rate difference, formula is $D \pm z_{1-\alpha/2} SE(D)$, D refers to point estimate of rate difference between two groups, z refers to standard normal distribution quantiles, two-sides α is 5%, $SE(D)$ is combined standard error,

$SE(D) = \sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}$, n_1 n_2 refer to sample size of both groups. Point and confidence interval estimate can be calculated using freq procedure with options of CMH in SAS software. However, when both group's rate are 100%, standard error can't be estimated using normal approximation method. And therefore confidence interval of rate difference can't be estimated.

Newcombe-Wilson Score Method

Newcombe-Wilson Score Method is firstly introduced by Newcombe in 1998. After comparing 11 methods of calculating incidence difference, Newcombe suggest Newcombe-Wilson Score is the best method to calculate the incidence difference from conservative and available aspect. It's recommended by FDA guideline to calculate the difference interval when both of treatment and controlled groups' rates are 0% or 100%.

Firstly, Wilson score method is used to calculate confidence intervals of single sample.

$(z^2_{1-\alpha/2} + n)\pi^2 - (z^2_{1-\alpha/2} + 2np)\pi + np^2 = 0$, π are lower and upper limit of confidence intervals, n is sample size of single group, p is effective rate of single group. So the formula of confidence intervals is

$(2np + z^2 \pm z\sqrt{(z^2 + 4npq)}) / 2(n + z^2)$, q is $(1-p)$. After obtain the confidence interval of each group, the confidence intervals of incidence difference can be calculated with the formula of

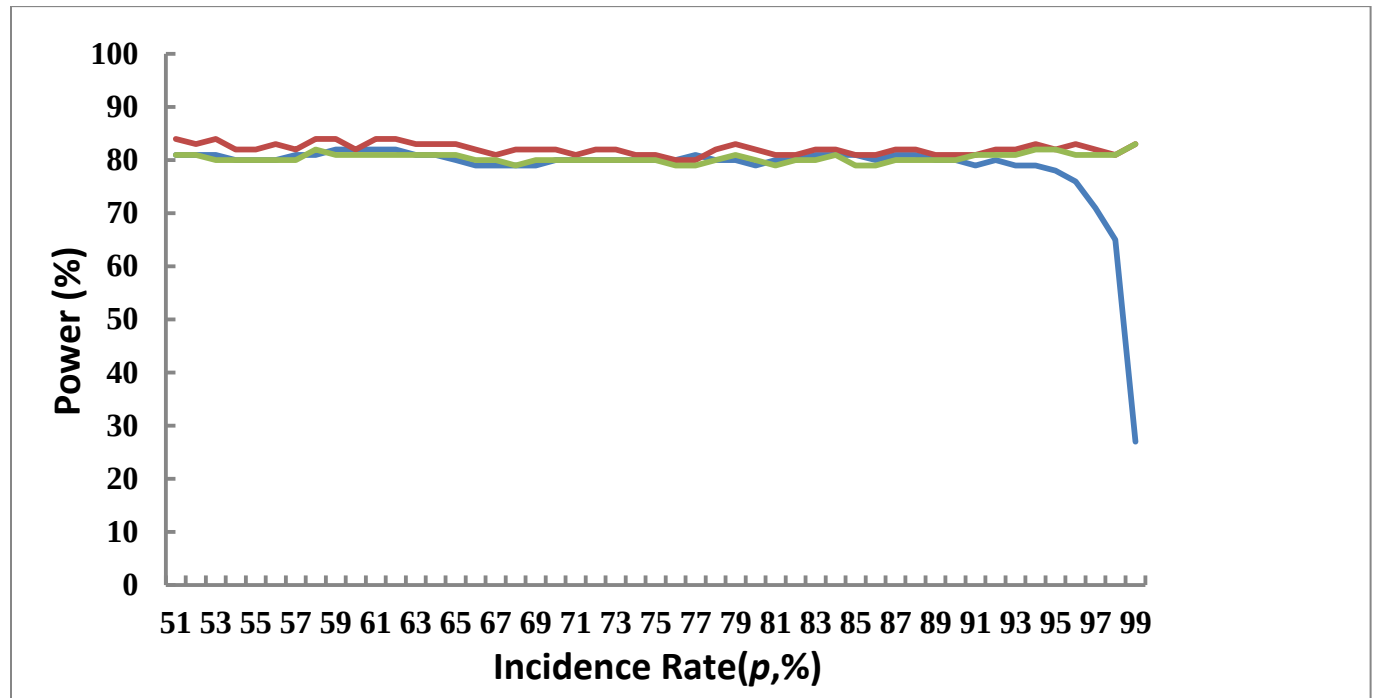
$D_l = D - \sqrt{(p_1 - l_1)^2 + (u_2 - l_2)^2}$, $D_u = D + \sqrt{(p_1 - l_1)^2 + (u_2 - l_2)^2}$, l_1 , u_1 , l_2 , u_2 refer to lower and upper limits of each group using Newcombe-Wilson score.

Continuous correction Newcombe-Wilson Score method is the most conservative of calculating the incidence difference, the formula is same as Newcombe-Wilson Score. The difference is in the calculation of single group's rate.

Formula is $(2np + z^2 + 1 \pm z\sqrt{(z^2 + 2 - 1/n + 4p(qn - 1)) / 2(n + z^2)})$. Since this method increase the width of confidence intervals, it Conservatively estimate rate difference between two groups.

The two methods can be used by programming in SAS or Excel, it's can be easily used.

RESULTS



From the figure displayed above about three methods, Normal approximation has power of more than 80% no matter how much is the incidence rate. Exact probability method also has similar power as normal approximation method, and have enough power using the sample size calculated by Normal approximation. However, if the event rate is more than 90%, Newcombe-Wilson Score Method will have insufficient power to calculate the incidence difference although it can estimate the confidence interval when both of the group's rates are 100%.

To avoid the disadvantage of Newcombe-Wilson Score Method, additional sample size estimate should be considered when both of groups' rates can be reach to 100% or 0% meanwhile.

**Table 1 Sample Sizes in Each Group Using Newcombe Wilson Score Method with
Different Non-inferiority Margins**

Margin (%)	SS (NO CC)	SS (CC)	Margin (%)	SS (NO CC)	SS (CC)
20	16	21	8.5	42	53
19	17	22	8	45	56
18	18	23	7.5	48	60
17	19	25	7	52	65
16	21	27	6.5	56	70
15	22	29	6	61	76
14	24	31	5.5	67	84
13	26	33	5	73	92
12	29	37	4.5	82	103
11	32	40	4	93	116
10	35	45	3.5	106	133
9.5	37	47	3	125	156
9	39	50			

NO CC: No continuity correction CC: No continuity correction

CONCLUSION

For the difference between independent proportions, three methods (Normal approximation, Exact Probability, Newcombe-Wilson Score) are presented,. Newcombe- Wilson Score method is recommended to deal with the case while both of two groups' rates are 0% or 100%, which is computationally very tractable irrespective of m and n, free from aberrations, and achieve better coverage properties than any except the most complex methods. However, additional sample size needs to be considered since this method won't have sufficient power when the trial meets the case both of two groups' rates are 0% or 100%.

REFERENCES

ACKNOWLEDGMENTS

RECOMMENDED READING

- Base SAS® Procedures Guide
- SAS® For Dummies®

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name:
Enterprise:
Address:
City, State ZIP:
Work Phone:
Fax:
E-mail:

<Paper title>, continued

Web:

Twitter:

The next two paragraphs are **required** and need to remain in the paper.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.