

# Stepwise Confidence Interval Method for MTD Studies with Binomial Populations

Na Yu<sup>1</sup>, Xiaoqing Tang<sup>1</sup>, Hanxing Wang<sup>2</sup>

<sup>1</sup>Department of Mathematics and Physics, Shanghai University of Electric Power, Shanghai, China

<sup>2</sup>School of Mathematics & Information, Shanghai Lixin University of Commerce, Shanghai, China

Email: mathyuna@163.com, tangxiaoqing5168@163.com

Received 2013

## Abstract

Now we extend one method into a sequence of binomial data, propose a stepwise confidence interval method for toxicity study, and also in our paper, two methods of constructing intervals for the risk difference are proposed. The first one is based on the well-known conditional confidence intervals for odds ratio, and the other one comes from Santner "small-sample confidence intervals for the difference of two success probabilities", and it produces exact intervals, through employing our method.

**Keywords:** Stepwise Confidence Interval; Practical Equivalence; Maximum Tolerated Dose

## 1. Introduction

We often face the maximum tolerated dose (MTD) evaluation of a new developed drug. Recently [1,2], Hsu and Berger proposed a stepwise confidence interval method for MTD studies and for dose-response study under equal variances [3]. But in fact, this is often in doubt. Recently, Tao *et al.* through employing the Stein's two-stage sampling method, proposes the stepwise confidence procedure for MTD evaluation and for identifying a minimum effective dose without any condition imposed on the variances [4]. But as we see in practice, the data we can collect are always discrete and are very limited.

And, currently, Tao *et al.* proposed a new stepwise confidence interval procedure to deal with the variance-free problem as a proper evaluation method [5]. By employing the Stern's two-stage sampling method, they achieve it. Now we extend one method into a sequence of binomial data [6], propose a stepwise confidence interval method for MTD study and can identify a minimum effective dose.

In their article they assume that a random sample  $Y_{i1}, Y_{i2}, Y_{i3}, \dots, Y_{in_i}$  is observed from the  $i$ th dose level, and considering the following one-way model

$$Y_{ij} = \mu_i + \varepsilon_{ij}, i = 0, 1, \dots, k+1; j = 1, 2, \dots, n_i$$

here  $\mu_0$  is the mean response of control group received a placebo and  $\mu_1, \mu_2, \dots, \mu_{k+1}$  are the mean responses to an increasing levels of exposure to a drug,  $\varepsilon_{ij}$  ( $i = 0, 1, \dots, k+1$ ) are *i.i.d.* normal variables with mean 0 and unknown variances  $\sigma_i^2$ .

However, as we can see the assumption of the population as normal distribution is quite unreasonable and often can bring us failure in making decision, for the data collected from the result of the experiment are often discrete, and even more the quantity is usually quite small due to lots of causes. Therefore, we could claim that their method is also not a mature method and can't always be reliable when we use it in practice, so a more widely reliable and useful method is necessitated.

Suppose that  $p_0$  is the response probability of the negative control group which receives placebo during the experiment, while  $p_{k+1}$  is the response probability of the positive control group.  $p_1, p_2, \dots, p_k$  is a series of response probabilities corresponding to a increasing level exposed to one drug, the dose level is denoted as  $d_1, d_2, \dots, d_k$ , which is typical in toxicological study and dose-response studies. And we think it is true: if the study fails to detect significant difference between the positive and the negative control groups, which are known to be different, then any lack of observed significant difference between a dose group and the negative control group may due to failed experimentation instead of closeness of their probabilities.

Assume an increasing order  $K$  levels dose of the new developed drug denote as  $d_1, d_2, d_3, \dots, d_k$  are allocated to  $K$  groups of people, and the  $p_1, p_2, p_3, \dots, p_k$  are the response probabilities corresponding to the dose levels. Suppose  $p_0$  is the response probability corresponding to the group which receives a placebo (negative control) during the comparisons, and similarly  $p_{k+1}$  are

the positive response probability.

So when  $\delta_L \leq p_i - p_0 \leq \delta_U$  hold, we should think it is denote effective when  $\delta_L \leq p_i - p_0$  and denote its maximum tolerated dose when  $p_i - p_0 \leq \delta_U$  as exporters assigned before trial.

## 2. Our Method

And each group satisfies

$$P(Y_i = m) = C_{ni}^m p_i^m (1 - p_i)^{ni-m},$$

where  $m = 1, 2, \dots, n_i, i = 1, 2, \dots, k + 1$ , and the  $p_i$  is parameter unknown.

To generalize Hsu and Berger's stepwise confidence interval procedure to the case of binomial population and to motivate our new stepwise, supposed  $(\Delta_l(X), \Delta_u(X))$  is the confidence interval with nominal level  $(1 - \alpha)\%$  for risk difference, and then let us rewrite the definition in Hsu and Berger.

**Definition 1.1** A confidence set,  $C(Y)$ , for  $\Theta$  is direct toward  $\Theta^*$  if, for every sample point  $y$ , either  $\Theta^* \subseteq C(Y)$ , or  $C(Y) \subseteq \Theta^*$ .

Then we have

**Lemma 1.1**  $(\Delta^{k+1}(X), 1)$  is a  $100(1 - \alpha)\%$  confidence interval for  $p_{k+1} - p_0$  direct toward  $\{p_{k+1} > p_0\}$

**Proof:** We have

$$P(p_{k+1} - p_0 \geq \Delta^{k+1}(X)) \geq 100(1 - \alpha)\% \\ \{p_{k+1} > p_0\} \equiv \{p_{k+1} - p_0 > 0\}$$

Then if we compare set  $(0, 1)$  with  $(\Delta^{k+1}(X), 1)$ , the result holds obviously.

**Lemma 1.2** For  $i = 1, 2, \dots, k$ , let

$$D_i^-(Y) = \min(\Delta_l^i(X), 0) \\ D_i^+(Y) = \max(\Delta_u^i(X), 0)$$

Then

$$D_i(Y) = \begin{cases} (D_i^-(Y), D_i^+(Y)) & D_i^-(Y) < 0 < D_i^+(Y) \\ [0, D_i^+(Y)) & D_i^-(Y) \equiv 0 \\ (D_i^-(Y), 0] & D_i^+(Y) \equiv 0 \end{cases}$$

is a  $100(1 - \alpha)\%$  confidence interval for risk difference  $p_i - p_0$ .

**Proof:** we have the relationship

$$D_i(Y) \supseteq (\Delta_l^i(X), \Delta_u^i(X))$$

And, so,

$$P(p_i - p_0 \in D_i(Y)) \geq 1 - \alpha$$

So  $D_i(Y)$  is a  $100(1 - \alpha)\%$  confidence interval for  $p_i - p_0$ .

**Lemma 1.3** For  $i = 1, 2 \dots k$ , let

$$C_i(Y) = \begin{cases} D_i(Y) & D_i(Y) \subseteq (\delta_l, \delta_u) \\ D_i(Y) \cup (\delta_l, \delta_u) & otherwise \end{cases}$$

And, then  $C_i(Y)$  is a  $100(1 - \alpha)\%$  confidence interval for  $p_i - p_0$  which direct toward  $(\delta_l, \delta_u)$ .

**Proof:** we have

If  $D_i(Y) \subseteq (\delta_l, \delta_u)$ , this implies that

$$P(p_i - p_0 \in D_i(Y)) \geq 1 - \alpha$$

If  $D_i(Y) \not\subseteq (\delta_l, \delta_u)$ , Since  $C_i(Y) = D_i(Y) \cup (\delta_l, \delta_u)$  We have

$$P(p_i - p_0 \in C_i(Y)) \geq 1 - \alpha$$

of course the set  $C_i(Y) \subseteq (\delta_l, \delta_u)$ , thus we finish the proof of this lemma.

## 3. Stepwise Procedure

If we arrange our experiment to an increasing dose of the new developed drug, and it can be answered the question  $\delta_L \leq p_i - p_0 \leq \delta_U$  in a stepwise fashion, continuing only when the answer is in an affirmative, until we find the first dose level whose toxicity response probability is not practical equivalent to the negative control group. In this paper we can also propose the stepwise confidence interval procedure for the binomial population as the following fashion,

**Step 0:**

If  $\Delta^{k+1}(X) > 0$ ,

Then we can assert the toxicity response probability between the positive control group and the negative is clinically difference, that is  $p_{k+1} - p_0 > 0$ , go to step 1.

Else assert  $p_{k+1} - p_0 > \Delta^{k+1}(X)$ , and then stop.

**Step 1:**

If  $D_1(Y) \subseteq (\delta_l, \delta_u)$ ,

We can assert that the toxicity response probability of the first dose level is practical equivalent to the negative control group, that is  $p_1 - p_0 \in (\delta_l, \delta_u)$ , go to step 2.

Else declare  $p_1 - p_0 \in C_1(Y)$ , and then stop.

.....

**Step k:**

If  $D_k(Y) \subseteq (\delta_l, \delta_u)$ ,

We can assert that the toxicity response probability of the first dose level is practical equivalent to the negative control group, that is  $p_k - p_0 \in (\delta_l, \delta_u)$ , go to step  $k + 1$ .

Else declare  $p_k - p_0 \in C_k(Y)$ , and then stop

**Step k + 1:**

Then we declare all the dose levels are safe, that is to say all the toxicity response probabilities are equivalent to the negative control groups and are tolerable.

To better understand the performance of our stepwise procedure, suppose step  $M$  ( $0 \leq M \leq k + 1$ ) is the step at which our stepwise procedure stops, which means the

subsequent comparison of more higher dose level is unnecessary, and find the maximum dose level whose toxicity response probability can be considered practical to the negative control group.

If  $M = 0$ , then we can declare that the sensitive of the experiment is not adequate, and a lower confidence is given to  $p_k - p_0$ ; if  $1 < M < k + 1$  then a confidence interval for  $p_M - p_0$  which contains  $(\delta_l, \delta_u)$  is given, also the confidence interval for  $p_i - p_0 \in (\delta_l, \delta_u)$ ,  $i = 1, \dots, k + 1$ , is given if  $M > 2$ ; finally, when  $M = k + 1$ , which means all the dose levels are safe, then the confidence interval for  $p_i - p_0$ ,  $i = 2, \dots, k + 1$ , which contained in  $(\delta_l, \delta_u)$  is also given.

We have the following theorem based on the procedure above:

**Theorem 1** Suppose our stepwise procedure stop at step  $M$  ( $0 \leq M \leq k + 1$ ), that is to say the toxicity response probability is practical equivalent to negative control group until the  $M$  dose level, and let  $C_i(Y)$  ( $i = 1, \dots, k + 1$ ) be the confidence interval define by Lemma 3.3,  $P_p$  {the incorrect decision we made until up to step  $M$ }  $\leq \alpha$ .

Note: when we perform our experiment in practice, a method should generate meaningful guarantee against incorrect decision, we can declare that the decision we have made is correct with probability higher than  $1 - \alpha$  which is pre-specified, thus we can control the family-I error rate so as to control the consumer's risk, we can express it in the form

$$P_p \left( \bigcap_{j=1}^M \{ \delta_l < p_j - p_0 < \delta_u \} \right) \geq 1 - \alpha$$

Our method generates a confidence interval for  $p_i - p_0$ , ( $i = 1, \dots, k + 1$ ) with coverage of pre-specified probability in a stepwise fashion, continuing our step until the first one confidence interval which does not contain the  $(\delta_l, \delta_u)$  is achieved and then we stop our procedure, whatever may the joint distribution be, we can infer that the incorrect decision we made is less than the nominal level which is pre-specified, that is to say the family I type error rate is well controlled according to the theorem.

#### 4. Application Example

Bovine growth hormone (bGH) is one of hormones which can promote cattle growth and milk secretion, and commonly it's low in cattle body. But add bGH to cattle feed which makes people worrying that these hormones are harmful. So it long became the focus of debate. To clarify these questions, Juskevich and Guyer (1990) have made many experiments, last report to FDA (Food and Drug Administration) pointed that experimental data did not show it's harmful when people drank the milk of

these cow feeding with bGH. Here we will analyze a set of data which is a group of these cow be feeding with bGH. This experiment includes four level dose (labeled as level 1 - 4) and a placeb (labeled as level 0) oral take and another group of positive control group (labeled as level 5). After 90 days of continuing experiment, mice's liver weight data as following **Table 1**:

When two doses' average response is equivalent, here significance  $\alpha < 0.05$ . And so, we have

$$D_1(Y) = (-1.62, 1.38), D_2(Y) = (-0.91, 2.10), \\ D_3(Y) = (-1.44, 1.56), D_4(Y) = (-1.86, 1.14).$$

Clearly, all  $D_i(Y)$  ( $i = 1, 2, 3, 4$ ) located in the interval of  $(\delta_l, \delta_u)$ , so actual equivalent be confirmed. This result is consistent with the expert's opinion. That is, such milk wouldn't do harm to people who drank it.

#### 5. Simulation Results

A computer simulation study was conducted to compare the behavior of our method with Dunnett method and DR method. We fixed that  $\alpha = 0.05, k = 5, n_i = 6, \mu_1 = 0, \frac{\sigma}{\sqrt{n}} = 1$ , and consider monotone dose response assumption.

We suppose that every three people as a group enter the test, every dose level use 2 groups of 6 people, observing their reaction to medication. We assume their toxicity probability have known distribution and every dose level is effective, when more than one people of the three are toxic, we will stop the experiment, and we will make another group take the same dose level. When the lower dose level completed, we will make a group of three people take the next upper dose level, and continue this process. When this process stops, we make the previous dose level as MED. Compare to the previous known MED dose level, we can know the error rate and the effectiveness of our method. We make the computer simulation 100 times for every case. Here following the computer simulation results (our method denoted by OM) as **Table 2**.

**Table 1. Effect of bGH on liver weight of mice ( $\bar{x} \pm s, n = 30$ ).**

| Groups        | Dose level | Delivery method | Dose (mg.kg <sup>-1</sup> ) | Liver weight |
|---------------|------------|-----------------|-----------------------------|--------------|
| Placebo group | 0          | po              | 0                           | 16.55 ± 3.00 |
|               | 1          | po              | 0.1                         | 15.61 ± 1.47 |
|               | 2          | Po              | 0.5                         | 15.74 ± 1.39 |
| bGH           | 3          | Po              | 5.0                         | 15.99 ± 1.94 |
|               | 4          | Po              | 50.0                        | 15.10 ± 2.21 |
|               | 5          | inj             | 1.0                         | 20.36 ± 2.22 |

**Table 2. Error rate of simulation of identify true MED.**

| Mean response   | Interval  | MED | OM   | Dunnett | DR   |
|-----------------|-----------|-----|------|---------|------|
| (0,1,2,3,4,5,6) | [2.5 3.5] | 4   | 0.23 | 0.31    | 0.25 |
|                 | [1.5 2.5] | 3   | 0.22 | 0.30    | 0.32 |
|                 | [0.5 1.5] | 2   | 0.28 | 0.28    | 0.27 |
| (0,0,0,2,2,4,6) | [1.5 2.5] | 5   | 0.18 | 0.20    | 0.21 |
|                 | [0.5 1.5] | 4   | 0.20 | 0.18    | 0.22 |
|                 | [0 1.5]   | 3   | 0.24 | 0.23    | 0.31 |
|                 | [0 1.0]   | 2   | 0.25 | 0.24    | 0.27 |
| (0,0,3,4,4,6,6) | [5.5 6.5] | 6   | 0.16 | 0.14    | 0.18 |
|                 | [4.5 5.5] | 5   | 0.20 | 0.22    | 0.30 |
|                 | [3.5 4.5] | 4   | 0.26 | 0.25    | 0.26 |
|                 | [2.5 3.5] | 3   | 0.12 | 0.15    | 0.14 |
| (0,0,0,5,5,6,6) | [4.5 5.5] | 5   | 0.38 | 0.40    | 0.31 |
|                 | [3.5 4.5] | 4   | 0.40 | 0.36    | 0.30 |
|                 | [2.5 3.5] | 3   | 0.33 | 0.34    | 0.35 |
|                 | [1.5 2.5] | 2   | 0.32 | 0.41    | 0.37 |
| (0,0,0,0,5,6,6) | [5.5 6.5] | 6   | 0.33 | 0.25    | 0.24 |
|                 | [4.5 5.5] | 5   | 0.29 | 0.22    | 0.23 |
|                 | [3.5 4.5] | 4   | 0.26 | 0.23    | 0.25 |
|                 | [2.5 3.5] | 3   | 0.44 | 0.42    | 0.46 |

We can see the simulation results from **Table 2**. First case, we assume the mean dose response of 5 drug experiment dose levels (including negative control group response which we set as 0 and positive control group which we set as 6, so there are 7 dose levels in total) are 1, 2, 3, 4, 5, when we set the interval is [2.5 3.5], if we assume the fourth dose level is the true MED and let computer randomly run 100 times according to certain rules, so the computer can automatically identify MED according to our method, then compare this MED with the pre-set true MED, we can get the error rate of our method, it is 0.23; while Dunnett method’s error rate is 0.31;DR method’s error rate is 0.25.

When we set the interval is [1.5 2.5], if we assume the third dose level is the true MED and let computer randomly run 100 times according to certain rules, we can see our method’s error rate , it is 0.22; while Dunnett method’s error rate is 0.30; DR method’s error rate is 0.32. when we set the interval is [0.5 1.5], if we assume the second dose level is the true MED and let computer randomly run 100 times, we know that our method’s er-

ror rate is 0.28; while Dunnett method’s error rate is 0.28; DR method’s error rate is 0.27. we also can see other cases simulation data from **Table 2**. we yet see case fourth and case fifth, when the situation is somehow extreme, for example, some dose levels’ responses are close to pre-set MED, or the deference of them is too big,

### 6. Conclusion and Further Discuss

From the examples, we can find that our method performs well. So our new method has much more confidence in practice. We hope that in the future, we will have further discuss about it, for example, with all sorts of dose response shape and then how to carry our method with previously designated confidents.

### 7. Acknowledgements

We should thank professor TAO Jian and this paper is supported by the National Natural Science Foundation of China (No: 60872060) and the Innovation Key Project of Shanghai Education Committee (No: 12ZZ193).

### References

- [1] J. C. Hsu and R. L. Berger, “Stepwise Confidence Intervals without Multiplicity Adjustment for Dose-Response and Toxicity Studies,” *Journal of the American Statistical Association*, Vol. 50, 1999, pp. 468-481.
- [2] J. Tao, N.-Z. Shi, J.-H. Guo and W. Gao, “Stepwise Procedures for the Identification of Minimum Effective Dose with Unknown Variances,” *Statistics & Probability Letters*, Vol. 57, 2002, pp. 121-131. [http://dx.doi.org/10.1016/S0167-7152\(01\)00197-3](http://dx.doi.org/10.1016/S0167-7152(01)00197-3)
- [3] J. Tao and N.-Z. Shi, “Stepwise Procedures under Unknown Variances for Toxicological Evaluation,” *Biometrical Journal*, Vol. 44, 2002, pp. 149-160. [http://dx.doi.org/10.1002/1521-4036\(200203\)44:2<149::AID-BIMJ149>3.0.CO;2-#](http://dx.doi.org/10.1002/1521-4036(200203)44:2<149::AID-BIMJ149>3.0.CO;2-#)
- [4] L. A. Hothorn, “Trend Tests in Epidemiology: P-Values or Confidence Intervals?” *Biometrical Journal*, Vol. 7, 1999, pp. 817-825. [http://dx.doi.org/10.1002/\(SICI\)1521-4036\(199911\)41:7<817::AID-BIMJ817>3.0.CO;2-C](http://dx.doi.org/10.1002/(SICI)1521-4036(199911)41:7<817::AID-BIMJ817>3.0.CO;2-C)
- [5] T. J. Santner and S. Yamagami, “Invariant Small Sample Confidence Intervals for the Difference of Two Success Probabilities,” *Communications in Statistics (Simulation and Computation)*, Vol. 22, 1993, pp. 33-59.
- [6] X.-Q. Tang and M.-C. Tan, “Stepwise Confidence Interval Method for Finding Drug’S MED in Trial,” *Journal of Shaoyang University*, No. 2, 2008, pp. 15-17.
- [7] X.-Q. Tang and H.-X. Wang, “Stepwise Confidence Interval Method for Finding Drug’s Minimum Effective Dose in Trial,” *Drug Evaluation Research*, Vol. 33, No. 1, 2010, pp. 58-62.