

**EXPRESS YOUR  
OPINION .....**

**Editor's Page**

Dear ICSA Members,

We added a list of some upcoming statistical meetings. In addition, the upcoming year 2001 applied statistics symposium and the 5<sup>th</sup> international conference of 2001 are sponsored by the ICSA. We'd like to notify you ahead of time and encourage your involvement by contacting the chairs of the corresponding events (details see announcement) or by sharing your research findings with the members. Your suggestions / comments are welcomed. Details inside ....

## Past Predicts Present and Future?

Statistics is born out from uncertainty, which plays an important role in many fields of knowledge. A recent counting and recounting of the US 2000 Presidential election ballots amounted to statistical error of margins quite critically. In this issue, the special topic of finance and the controversial statistical issue in active controlled clinical trials will broaden our minds on the role of uncertainty.

People tend to make use of the existing data to predict the immediate future. But, could it be present in a broad sense? As evidenced by recent Dow Jones and Nasdaq statistics, financial markets are ever more volatile and changing over time. The Futures Market research is vital. Model justification, taking into account random fluctuations, requires the use of statistical methods for testing its appropriateness. Statistical tools frequently used in finance literature of time series, stochastic volatility and Bayesian are discussed with their motivations behind.

The drug approval process in the United States is in place to assure that drugs available to the American public are effective and safe. As more and more drugs are approved, it may no longer be ethical to continue the practice of placebo-controlled studies. A new drug may be compared to aspirin to test for its effectiveness in preventing or treating thrombolytic and stroke patients. This type of trials is the so-called active-controlled studies. Those trials, which showed that aspirin is effective, are the placebo-controlled trials.

In the column of Controversial Statistical Issue, several serious, however, entertaining articles, e.g., "Non-inferiority: A Dangerous Toy?", depict philosophical dilemma and conceptual difficulties in conducting active-controlled studies for showing that the new treatment is not unacceptably worse than the active control agent by some pre-specified margin. The trick is to link the new treatment from the present trial with the placebo in the past trial. The big question is can such a "not unacceptably worse" claim lead one to infer that the new treatment would have been better than no treatment or the placebo had the placebo been in the present active-controlled trial with reasonable scientific evidence?

As the number of our contributing writers increases by the issue, it is as exciting as ever to put together the work of statisticians from the US, Asia and Europe. Some statistics organizations are showing enthusiasm in our Bulletin. To build on this, we are happy to exchange the meeting announcements with other organizations, e.g., the Society for Clinical Trials, to publicize our organization. We begin by publishing a meeting calendar in this issue. We hope you enjoy this issue.

*Sue-Jane Wang*  
**Editor-in-Chief**

**IN THIS ISSUE ...**

Messages .....	2
Minutes .....	3
Reports .....	6
Keynote Speech .....	7
Professor P.C. Tang .....	13
Special topic–Finance .....	16
Controversial Statistical Issue – Active-Controlled Clinical Trials .....	25
Announcements .....	41
Regional Activities .....	50
Statistics' Delight / 統計趣聞	51
Members/Activities .....	53
Advertisement .....	55
Financial Reports .....	57
Memberships Form/Info ..	60

**GET INVOLVED IN  
NEXT ISSUE .....**

Special topic – Actuarial  
Statistics' Delight/ 統計趣聞  
Controversial statistical  
Issue – Bootstrapping vs.  
Markov Chain Monte Carlo

**Get involved in the next  
issue by email your articles  
to the Editorial Board at  
[WANGS@CDER.FDA.GOV](mailto:WANGS@CDER.FDA.GOV)**



# **!!! Controversial !!! Statistical Issue**

Extracted from ICSA Bulletin January 2001 issue, pp 25-40.  
Editor: Sue-Jane Wang, Ph.D.

## **Active Controlled Clinical Trials**

### **Non-Inferiority Trials: Does Sloppiness Bias Toward No Difference?**

By Irving K. Hwang, Ph.D.  
Harvard Clinical Research Institute

“Assay Sensitivity,” as defined in ICH E10 Guidance [1], is a property of a clinical trial that has the ability to distinguish an effective treatment from a less effective or ineffective treatment. In a superiority trial, when a test treatment is shown superior to the control (placebo or active control), the finding itself demonstrates assay sensitivity. However, in a trial intended to show a test treatment is non-inferior to an active control, assay sensitivity can only be deduced from historical evidence of sensitivity to drug effects of the chosen

active control and appropriate trial conduct (high quality) of the current non-inferiority trial.

It is well understood that in a superiority trial with an intention to show difference between treatments, there is strong imperative to design and conduct the trial with good quality to increase the likelihood of demonstrating assay sensitivity. Whereas, in a trial intended to show no difference of a particular size (i.e., a non-inferiority margin) between treatments, there is a much weaker stimulus to optimize study design and reduce study errors due to the expectations of patients and investigators on receiving active treatments. It is generally believed that in a non-inferiority trial, trial sloppiness (poor trial quality) would bias toward no difference [2,3] and in turn, increase the likelihood that an ineffective treatment could be found non-inferior. Is this belief statistically plausible?

Prior to answering this question, it is necessary to compare the null and alternative hypotheses of the

superiority trial with those of the non-inferiority trial. When the concept of hypothesis testing, or

equivalently, the notion of confidence interval, is fully appreciated, the similarity between these two types of trials, as shown below, will become obvious.

Superiority Trial	Non-Inferiority Trial
<p><b>H<sub>0</sub></b> (no treatment difference):  <math display="block">\mu_t - \mu_c = 0</math>                     vs.  <b>H<sub>1</sub></b> (test treatment better):  <math display="block">\mu_t - \mu_c &gt; 0.</math>                     To reject the null hypothesis (showing test treatment superior), it is necessary that  <math display="block">z = (\bar{x}_t - \bar{x}_c) / s(\bar{x}_t - \bar{x}_c) &gt; 1.96.*</math></p>	<p><b>H<sub>0</sub></b> (test treatment inferior):  <math display="block">\mu_t - \mu_c &lt; -\delta</math>                     vs.  <b>H<sub>1</sub></b> (test treatment non-inferior):  <math display="block">\mu_t - \mu_c \geq -\delta.</math>                     To reject the null hypothesis (demonstrating test treatment non-inferior within the margin, <math>\delta</math>), it is necessary that  <math display="block">z = ((\bar{x}_t - \bar{x}_c) + \delta) / s(\bar{x}_t - \bar{x}_c) &gt; 1.96.*</math></p>

\* Assuming the test is one-sided at  $\alpha = 0.025$  level with a critical value = 1.96. Where  $\mu_t$  and  $\mu_c$  represent the population means and  $\bar{x}_t$  and  $\bar{x}_c$  the observed (sample) means for the test and control treatments, respectively,  $s(\bar{x}_t - \bar{x}_c)$  represents the observed (sample) standard error of the mean difference, and  $d = \bar{x}_t - \bar{x}_c$  such that  $s(\bar{x}_t - \bar{x}_c) = s_d$

In a superiority setting, it needs a large numerator (large observed mean difference,  $d$ ), and/or a small denominator (small observed standard error of the mean difference,  $s_d$ ), to make the test statistic significant ( $z > 1.96$ ,  $p < 0.025$ ), and thus show test treatment superior. Similarly, in non-inferiority setting, it also needs a large positive numerator,  $(d + \delta)$ , and/or concurrently a small denominator,  $s_d$ , to reach significance, and thus demonstrates test treatment non-inferior. In fact, good trial design and conduct with large observed differences and/or small standard errors are necessary regardless whether a trial is designed to show superiority or non-inferiority.

Now, let us examine whether trial sloppiness in a non-inferiority trial indeed biases toward no difference and in turn, increases the likelihood that an ineffective treatment could be found non-inferior. In fact, the test statistic for the non-inferiority trial is:

$$z = (d + \delta) / S_d$$

Reject **H<sub>0</sub>**:  $\mu_t - \mu_c < -\delta$  to conclude non-inferiority of test to active control, if  $z > 1.96$ . Equivalently, one can use the confidence interval (CI) approach to claim non-inferiority, if,  $[C_L, \infty)$ , the 1-sided 97.5% CI for  $\mu_t - \mu_c$ , is included in  $[-\delta, \infty)$  or  $-\delta < C_L$ , where  $C_L$  is the lower limit of the 1-sided 97.5% CI.

No difference means  $d = 0$ . Whether one can reject **H<sub>0</sub>** and claim non-inferiority will depend on the relative magnitude of  $\delta$  to  $s_d$  (i.e. the ratio of  $\delta / s_d$ ) when  $d = 0$ . If the ratio of  $\delta / s_d$  is greater than 1.96, then one can reject **H<sub>0</sub>** and claim non-inferiority. Otherwise, one would have to declare inferiority, since  $\delta$  is the appropriately chosen non-inferiority margin which is usually small unless the standard error of  $d$  is much smaller. Therefore, no difference (i.e.,  $d = 0$ ) does not necessarily imply non-inferiority.

Sloppiness usually introduces noise. Noise implies increased variability with wider confidence interval and in turn it biases toward  $H_0$ . In addition, sloppiness generally introduces bias. Bias may happen in either direction (i.e., reduced or increased observed difference, d). Therefore, sloppiness may bias toward either  $H_0: \mu_t - \mu_c < -\delta$  (inferiority) or  $H_1: \mu_t - \mu_c \geq -\delta$  (non-inferiority), though the latter is the major concern in

conducting non-inferiority trials.

Discussions on this issue can also be found in Hwang & Morikawa [4] and Hauck & Anderson [5].

#### References:

1. ICH E10 Guidance. Choice of control group and related design issues in clinical trials. Step 4 Document, July 11, 2000.
2. Temple, R. Government viewpoint of clinical trials. *Drug Inf J.* 1982; 16:10-17.
3. Temple, R. Clinical trials in seriously ill patients – design considerations. *PAACNOTES.* 1990; 133-143.
4. Hwang, I.K., Morikawa, T. Design issues in non-inferiority/equivalence trials. *Drug Inf J.* 1999;33 (4):1205-1218.
5. Hauck W.W., Anderson, S. Some issues in the design and analysis of equivalence trails. *Drug Inf J.* 1999;33 (1):109-118.

## Non-inferiority: A Dangerous Toy?\*

By H.M. James Hung, Ph.D.

Division of Biometrics I, Office of Biostatistics, CDER, FDA, Rockville, MD, USA

A clinical trial as an experiment on humans often needs to be run using a complex design for ethical reasons, that is, the welfare of patients in and out of the trial must be properly considered. As more or more effective treatments have been available to the patients, the clinical trial for studying the effectiveness of an experimental treatment, particularly in the same class of the effective medical products, can be very difficult to design. At least, use of placebo can be challenged. The only viable choice seems to be use of a positive control. A difficult task is choice of an effective treatment from the historical trial experiences to serve as a control for the experimental treatment to compare against in the positive control

trial. If it can be shown more effective than the positive control, the new treatment is usually accepted as an effective treatment.

Problems arise when a greater effect with the new treatment may not be possible to show, perhaps because in truth the effect difference is minimal between the new treatment and the control. Under this scenario, the effectiveness of the new treatment will likely be established by arguing a minimal difference in the treatment effect or that the new treatment is minimally less effective than the control. The former is the concept of equivalence and the latter is non-inferiority, which is pursued more often

than equivalence in therapeutic clinical trials. In the first place, there seems to be no problem with the non-inferiority setting. All needed is setting up a non-inferiority margin that is related to the acceptable degree of difference in treatment effect. It can be subjectively and arbitrarily defined. Certainly, such a way of selection causes a lot of concerns because of no reference to the position of the invisible placebo. So, recently, many scientists argue that choice of the non-inferiority margin should take into consideration the control effect, which is at best estimated from some historical trials. What a noble argument! It opens cans of worms. How should one do meta-analyses (or integrated analyses or pooled, combined,...) to estimate the control effect? By random-effect or fixed-effect modeling? Should we incorporate the error of the estimate in assessing the type I or II error probabilities incurring in non-inferiority statistical testing? Oh! No! this involves cross-study comparisons. In order to interpret the results of such comparisons, we require that both trial populations are randomly sampled from the same population (we know that random sampling is never in place) or well representative of the target population (we know that this is always not verifiable). Such comparisons are very problematic at least! In fact, talking about type I error probability for non-inferiority testing is wrong! But, we can calculate a p-value for comparing the new treatment with the invisible placebo and it is  $10^{-16}$ . No! Posterior probability will work! Why isn't type I error an issue as it always is in any clinical experiment? We need a placebo arm but Hum! This is a moot point....

Let us take a step back. What is the non-inferiority hypothesis? For ease of presentation, let me use alphabetical letter

to represent both a treatment group and its expectation of an interested response. Let T-C be the effect of the new treatment T relative to the concurrent positive control C. From the historical data, we have the parameters  $C_0$  and  $P_0$  of the intended-to-use positive control and the placebo, respectively. Originally, one possible null hypothesis for non-inferiority is

$H_0: T-C \leq \delta$ , where  $\delta$  is the non-inferiority margin. This margin is only vaguely defined by referring to the estimated  $C_0$  and  $P_0$  at best. Thus, it does not make sense to write  $\delta = f(C_0-P_0)$  for some function  $f$  because the essential questions of interest apply only to the current trial population. In addition, if this expression is sensible, then there is never a problem with statistical inference because the test statistic "carefully" constructed using approximately unbiased estimators for both sides of the inequality of  $H_0$  will have an approximately correct type I error probability associated with this null hypothesis. Can  $\delta$  be some function of C-P (P is the corresponding parameter of the invisible placebo that does not exist in the concurrent positive-control study)? This is exactly suggested by the essential questions of interest (the patients with the current disease conditions are of greater interest, not the historical patient population). Because of missing P, one may argue that the type I error probability is not computable. However, if one believes that C-P is estimable from the historical trials, then this hypothesis is testable. Moreover, one can assess the impact of the error in estimation of C-P using the estimate of  $C_0-P_0$  on the error probability in testing  $H_0$ ; see the reference [1].

Averaging is a well-known powerful tool in the human inventions. Let us take a simple average of C-P and  $C_0-P_0$ . The resulting average reflects the center of the

two centers. The deviations between the two centers and the resulting average describe heterogeneity between the centers. Imagine that there are many future positive control trials using the same positive control to develop new medicine. Continuous application of averaging will construct an interesting center and descriptor for heterogeneity between the centers, i.e., the parameters C-P from the historical trials and future positive control trials. This is, in essence, the basis of Bayesian framework. In some sense, such an averaging process updates the location of the control effect C-P from the past to the future. This concept is very natural. But wait a minute. There will never be any future data for C-P and the only data we have for C-P is from the estimates of  $C_0-P_0$ . While updating the location of C-P, the data remains at the estimates of  $C_0-P_0$ . So why will this

framework have more advantage in terms of resolving the controversy?

In brief, we realized the controversy long time ago and got away from it for the most part. Now, we are part of the controversy in playing the dangerous toy – NON-INFERIORITY. Is it really that dangerous?

#### Reference

[1] Wang, S.J., Hung, H.M.J. and Tsong, Yi (2000), "Utility and pitfall of some statistical methods in active controlled clinical trials" (submitted for peer review).

\*The views expressed in this article do not necessarily represent those of the U.S. Food and Drug Administration. This work was partially supported by RSR fund #RSR 01-20 of Center of Drug Evaluation and Research, Food and Drug Administration.

## **Alternatives for Discounting Historical Data in the Analysis of Non-Inferiority Trials**

By Steven M. Snapinn

Merck Research Labs, West Point PA 19486

### INTRODUCTION

Non-inferiority trials are clinical trials in which the goal is to demonstrate that the effect of an experimental treatment is not inferior to that of an active control by more than a specific margin. Often, a secondary goal is to indirectly compare the experimental treatment

with placebo. The most common situation where this occurs is when the active control has been previously studied versus placebo, but including a placebo in a new trial would be considered unethical. In some cases this indirect evidence for an effect versus placebo has been used as evidence for regulatory approval of a new therapy.

The indirect comparison of the experimental treatment with placebo depends for its validity on two key assumptions: that the effect of the active control in the current trial is the same as it was in the historical placebo-controlled trials (constancy), and that the current trial has the same ability to distinguish the active control

from placebo as the historical active-controlled trials (assay sensitivity). It is disconcerting that these assumptions can not be directly verified within the current trial and that there are often reasons to doubt their truth.

These assumptions are discussed extensively in ICH E-10 guideline [1] and elsewhere [2-5]. Because of the questions surrounding the validity of the indirect comparison of the experimental treatment with placebo, it would seem reasonable to apply some type of discounting to the historical placebo-controlled data; that is, to give them less weight than if they were an integral part of the current trial. The purposes of this paper are to further discuss the concept of discounting, to show that various methods of analyzing non-inferiority trials can be put into the common context of discounting, and to compare the degree of discounting associated with these methods.

## METHODS

Suppose the analysis variable of interest is binary, and that the method of analysis will be based on the log odds ratio, and that the summary results of the

historical placebo-controlled trials and the results of the current trial are denoted as follows:

$B_{SP}, V_{SP}$  the log odds ratio and its variance for the active control (standard) relative to placebo (from the historical trials).

$B_{XS}, V_{XS}$  the log odds ratio and its variance for the experimental treatment to the active control (from the current trial).

If we can assume constancy and assay sensitivity and do not want to discount the historical data, then it is relatively simple to indirectly compare the experimental treatment to placebo, while fully accounting for random variation in the effects of all treatments. We simply sum the log odds ratios for the experimental treatment relative to the active control and for the active control relative to placebo to obtain an approximately normally distributed random variable with variance equal to sum of the two variance terms:  $Z_0 \sim N(B_{XS} + B_{SP}, V_{XS} + V_{SP})$ .

### *Discounting Approach #1: Setting a Non-Inferiority Margin*

A frequent method of analyzing non-inferiority trials is to first specify a non-inferiority margin, denoted by  $\delta$ , and then to compare the results of the current trial with that margin; if the confidence interval for the effect of the experimental treatment relative to the active control is entirely above this margin, then experimental treatment is declared non-inferior to the active control. In order to draw an indirect inference relative to placebo, the non-inferiority margin can be selected based on the historical placebo-controlled results. While it is not always recognized, this approach is extremely inefficient relative to the simple approach defined above, and therefore in effect includes a form of discounting.

Suppose, for example, that the non-inferiority margin is set to be the negative of the lower bound of the 2-sided 95% confidence interval for the effect of the active control relative to placebo in the historical trials, or  $\delta = -(B_{SP} - 1.96\sqrt{V_{SP}})$ , and that non-inferiority will be declared if the lower bound of the 2-sided 95% confidence interval for the effect of the experimental

treatment relative to the active control in the current trial is greater than  $\delta$ , or  $B_{XS} - 1.96\sqrt{V_{XS}} > \delta$ .

This approach is essentially equivalent to indirectly comparing the experimental treatment to placebo using the following test statistic, in that a significant difference based on the test statistic corresponds to meeting the criterion involving the non-inferiority margin:  $Z_1 \sim N(B_{XS} + B_{SP}, (\sqrt{V_{XS}} + \sqrt{V_{SP}})^2)$ . Note that  $Z_0$  and  $Z_1$  are similar, except that rather than pooling variances,  $Z_1$  discounts by pooling standard errors.

*Discounting Approach #2: Preserving a Fraction of the Active Control's Effect*

The approaches described thus far are intended to indirectly compare the experimental treatment to placebo and to declare it effective if it has any benefit over placebo, regardless of the relative benefits of the experimental treatment and the active control. In some cases the goal involves a higher standard, at least on the surface: In order for an experimental treatment to be declared effective it must not only be superior to placebo, but also must preserve a specific fraction

of the active control's effect relative to placebo. For example, in order to demonstrate preservation of half of the active control's effect, one must rule out an effect of the experimental treatment relative to placebo of less than  $\frac{1}{2}B_{SP}$ .

There are two ways to view this requirement. On its surface, this is a new standard of effectiveness for the experimental treatment. This is somewhat controversial, since it is inconsistent with the usual standard for placebo-controlled trials. However, another less controversial viewpoint is that this is simply another means of discounting the historical data. In this view, a positive result would not necessarily mean that the experimental treatment preserves any specific portion of the active control's effect, but it would give greater confidence that the experimental treatment is superior to placebo.

While there are more sophisticated approaches that could be used [8], a very simple statistic that illustrates the discounting associated with preservation of 50% of the active control's effect is as follows:  $Z_2 \sim N(B_{XS} + \frac{1}{2}B_{SP}, V_{XS} + \frac{1}{4}V_{SP})$ .

*Discounting Approach #3: Double-Discounting*

The final approach discussed here is referred to as the double-discounting approach, and has been proposed by the FDA in at least one clinical setting [9]. In this approach one applies the preservation of 50% of the active control's effect to the non-inferiority margin, or  $\delta = -\frac{1}{2}(B_{SP} - 1.96\sqrt{V_{SP}})$ .

This approach combines the discounting associated with setting a non-inferiority margin with the discounting associated with preservation of 50% of the active control's effect. Using the same format as before, a simple test statistic associated with this approach is given here:  $Z_3 \sim N(B_{XS} + \frac{1}{2}B_{SP}, (\sqrt{V_{XS}} + \sqrt{\frac{1}{4}V_{SP}})^2)$

A COMPARISON OF APPROACHES

Table 1 compares the statistics produced by  $Z_0$ ,  $Z_1$ ,  $Z_2$  and  $Z_3$  for various combinations of  $B_{SP}$ ,  $B_{XS}$ ,  $V_{SP}$  and  $V_{XS}$ . All else held constant, all statistics increased,

indicating increasing evidence that the experimental treatment is superior to placebo, as either the strength of evidence that the active control is superior to placebo increased (*i.e.*, as  $B_{SP}$  increased for a fixed  $V_{SP}$ ), or as the strength of evidence that the experimental treatment is superior to the active control increased (*i.e.*, as  $B_{XS}$  increased for a fixed  $V_{XS}$ ). In all situations studied the statistics produced by  $Z_1$ ,  $Z_2$  and  $Z_3$  were smaller than the statistic produced by  $Z_0$ , and the statistic produced by  $Z_3$  was as small as or smaller than all others.  $Z_1$  tended to be larger than  $Z_2$ , but not uniformly. The factors that tended to increase  $Z_2$  relative to  $Z_1$  were a relatively small level of evidence that the active control is superior to placebo and a relatively large level of evidence that the experimental treatment is superior to the active control.

Note that the degree of discounting produced by these statistics was sometimes fairly extreme. Take as an example the next-to-last row of the table. The statistic produced by  $Z_0$  was 3.578, which corresponds to a highly significant 2-sided

p-value of 0.0003, while the p-values corresponding to  $Z_1$ ,  $Z_2$  and  $Z_3$ , respectively, were 0.0077, 0.0524 and 0.1096.

## CONCLUSIONS

It is clearly appropriate to discount the historical placebo-controlled data to some degree when indirectly comparing an experimental treatment to placebo following a non-inferiority trial, due to the uncertainty surrounding the assumptions that form the basis for the validity of that comparison. However, the level of discounting required is less clear since it is somewhat subjective and may be different in different situations. As described in this paper, finding the right level of discounting may be complicated by a lack of recognition of the sources of discounting associated with common methods of analysis of non-inferiority trials. It could be argued that we do not yet have a logical and consistent approach for discounting, and it may turn out that neither specification of a non-inferiority margin nor preservation of a fraction of the active control's effect will be the preferred method of discounting in the future. For example, applying a multiplicative constant to the variance for the effect of the active control relative to placebo from the historical trials has a logical appeal and may turn out to be the method of choice. Hopefully, the conceptual framework described here will be helpful in approaching this problem.

## REFERENCES

1. ICH E-10 Guideline: Choice of control group in clinical trials. Released for Consultation at *Step 2* of the ICH Process on 7 May 1999.
2. Snapinn, SM. Non-inferiority trials. *Current Controlled Trials in Cardiovascular Medicine* 2000, 1:19-21.
3. Temple R, Ellenberg SS. Placebo-controlled trials and active-controlled trials in the evaluation of new treatments. Part 1: Ethical and scientific issues. *Annals of Internal Medicine* 2000, 133:455-463.
4. Ellenberg SS, Temple R. Placebo-controlled trials and active-controlled trials in the evaluation of new treatments. Part 2: Practical issues and specific cases. *Annals of Internal Medicine* 2000, 133:464-470.
5. Siegel JP. Equivalence and non-inferiority trials. *American Heart Journal* 2000, 139:S166-S170.

6. Hasselblad V, Kong DF. Statistical methods for comparison to placebo in active-controlled trials. To appear in *Drug Information Journal* 2001.
7. CBER/FDA Memorandum. Summary of CBER considerations on selected aspects of active controlled trial design and analysis for the evaluation of thrombolytics in acute MI 1999.

**Table 1**  
Statistics Produced by  $Z_0, Z_1, Z_2$  and  $Z_3$  for Various Combinations of  $B_{SP}, B_{XS}, V_{SP}$  and  $V_{XS}$

$B_{SP}$	$B_{XS}$	$V_{SP}$	$V_{XS}$	$Z_0$	$Z_1$	$Z_2$	$Z_3$
2	-1	1	1	0.707	0.500	0.000	0.000
2	0	1	1	1.414	1.000	0.894	0.667
2	1	1	1	2.121	1.500	1.789	1.333
4	-1	1	1	2.121	1.500	0.894	0.667
4	0	1	1	2.828	2.000	1.789	1.333
4	1	1	1	3.536	2.500	2.683	2.000
2	-1	½	1	0.816	0.586	0.000	0.000
2	0	½	1	1.633	1.172	0.943	0.739
2	1	½	1	2.449	1.757	1.886	1.478
4	-1	½	1	2.449	1.757	0.943	0.739
4	0	½	1	3.266	2.343	1.886	1.478
4	1	½	1	4.082	2.929	2.828	2.216
2	-1	¼	1	0.894	0.667	0.000	0.000
2	0	¼	1	1.789	1.333	0.970	0.800
2	1	¼	1	2.683	2.000	1.940	1.600
4	-1	¼	1	2.683	2.000	0.970	0.800
4	0	¼	1	3.578	2.667	1.940	1.600
4	1	¼	1	4.472	3.333	2.910	2.400

## Equivalence studies can not be used to claim equivalence of two active treatments

Anders Källén, Ph.D., M.D. & Per Larsson, Ph.D.  
AstraZeneca R&D Lund, Lund, Sweden

In the ICH guidelines on Choice of Control Group and Related Issues in Clinical Trials (1), the concepts "equivalence trial" and "non-inferiority trial" are used to denote a clinical trial designed to demonstrate efficacy of a new drug by showing that it is similar in efficacy to a standard agent. As discussed in the guidelines, equivalence trials and non-inferiority trials may be an alternative to placebo controlled trials for showing efficacy of a new drug. The names "equivalence" and "non-inferiority"

however indicate that these trials could be used for far stronger claims than efficacy, namely to claim that a new treatment is "as effective as" (equivalent to) or "at least as effective as"

(non-inferior to) a comparator treatment. We want to argue that such use of equivalence and non-inferiority trials is incorrect, and we propose an alternative way to address the question of therapeutic equivalence.

To illustrate the problems we will consider the following example. Assume that we want to prove that 200 µg of the glucocorticosteroid A has the same clinical effect on asthma as 400 µg of the glucocorticosteroid B. To use a therapeutic equivalence trial, we first need to pre-define our equivalence limits. One suggestion for this (2) is that the mean difference in morning Peak Expiratory Flow should be within -15 to 15 L/min (i.e. a 95% confidence interval for the mean difference should be entirely within this interval). If we run a trial and get the 95% confidence interval -14 to 12 L/min, for example, can we then draw the conclusion that the two treatments are equivalent?

As a statistical procedure, the approach can be regarded as valid and the conclusion of equivalence is correct. From a medical point of view we would argue that the approach can be criticized and that the conclusion is not always correct. The first objection regards the pre-defined equivalence limits. If it is generally accepted that a mean difference less than 15 L/min is clinically irrelevant, there is no problem, but if there is disagreement about this, the conclusion of equivalence must be questioned. Anyone believing that smaller mean differences than 15 L/min are clinically important have the right to argue that the study is inconclusive. The fact that the equivalence limits are pre-specified does not make them generally accepted. Our first conclusion is therefore that the pre-specification of equivalence limits does not add any value to an equivalence study. The important result should be the actual confidence limits obtained in the study – the clinical interpretation of these are then what decides whether one could consider the difference small enough to be clinically irrelevant.

Our second objection regards the ability of the study to detect possible differences between the treatments. The ICH guideline clearly states that for an equivalence or non-inferiority trial to be valid, it must be supported with evidence that the trial had ability to distinguish an effective treatment from a less effective or ineffective treatment (so called assay sensitivity). This is even more important for a study aiming to conclude equivalence of two active treatments. In the example above, assume that we instead

compared 200 and 400 µg of the same steroid. If we then got a confidence interval contained within the equivalence limits, should we then conclude that these two doses of the steroid are equivalent? If we accept the equivalence limits, it is probably correct that for these patients there is, on average, no point in increasing the dose. This does however not necessarily apply to a different patient population and generalization of the conclusion to a wider population than the actual study population is probably incorrect.

The same objection applies to the study comparing 200 µg of steroid A with 400 µg of steroid B. The strongest conclusion we should be allowed to make is that in the actual study population, the treatments produce on average the same result. Many observers would unfortunately draw one further conclusion – that steroid A is twice as potent as steroid B. This conclusion can not be drawn since there is nothing in the results to show that not also 200 and 400 µg of steroid A would have been considered

therapeutically equivalent in this study population. To conclude that there is a difference in potency there must be evidence that the study had ability to distinguish between different doses of the steroids.

Based on our two objections we suggest a different approach to therapeutic equivalence, and in fact, that the concept of therapeutic equivalence is not needed at all.

The approach we suggest is restricted to the situation when we have two treatments that can be given in different doses (the doses do not have to be approved or marketed). We then suggest a study where two doses of treatment A are compared to one dose of treatment B. We need to choose our patients and design the study so that we can statistically demonstrate a difference between the two doses of treatment A. In our previous example, we choose 100 and 400 µg of steroid A and 200 µg of steroid B. If the two doses of steroid A differ statistically significantly we approximate the dose response curve for steroid A with a linear function of log-dose. From this approximation we estimate the dose of steroid A corresponding to 200 µg of steroid B and calculate a confidence interval for the estimate (2,3). Assume that we get the estimate 186 µg with confidence interval 120 to 270 µg. If we are content with working with dose doubling steps, the interpretation could be that 200 µg of steroid A is therapeutically equivalent with 400 µg of steroid B. The argument would be that half the dose of steroid A, 100 µg, is too low and twice the dose, 400 µg, is too high.

If we are not content with working only with dose doubling steps, more specific equivalence limits must be agreed upon. The fact that we now work on the dose scale may simplify this. We would however argue that the important result still is the confidence interval for the dose of steroid A corresponding to 200 µg of steroid B and, in fact, that there is no need for a definition of therapeutic equivalence. By providing the estimate for the equivalent dose and confidence limits for this, it should be possible for the reader to assess to what extent this means therapeutic equivalence in practical terms. Whether the reader is the doctor, contemplating switching treatment for his patients,

or the regulatory reviewer who must assess whether or not the evidence submitted is sufficient for a specific claim, the information should be sufficient for a decision to be made.

#### References:

1. ICH Harmonised Tripartite Guideline, Choice of Control Group and Related Issues in Clinical Trials, Recommended for adoption at step 4 of the ICH process on July 2000 by the ICH steering committee.
2. Boulet, Cowie, Johnston, Krakovsky, Mark . Comparison of Diskus™ inhaler, a new multidose powder inhaler, with Diskhaler™ inhaler for the delivery of salmeterol to asthmatic patients. *Journal of Asthma*, 32(6), 429 – 436 (1995).
3. Källén, Larsson. Dose response studies: How do we make them conclusive? *Statistics in Medicine* 1999;18:629 – 641.
4. Källén, Larsson. On the definition of therapeutic equivalence. *Drug Information Journal*, Vol. 34, pp. 349 – 354, 2000.

# Placebo Control, Historical Control and Active Control Trials\*

By Yi Tsong, Sue-Jane Wang, Lu Cui, James H.M. Hung  
Office of Biostatistics, Center of Drug Evaluation and Research, FDA

In order to demonstrate that a test treatment is effective, one needs to conduct a clinical trial to test for

$$\begin{aligned} H_0: T = R \text{ versus} \\ H_A: T > R, \end{aligned}$$

where R is the reference treatment used as control. The test treatment is shown to be effective when  $H_0$  is rejected (often required at  $\alpha=0.025$  level). The most common design for confirmatory clinical trials is the parallel group design. Based on the

different types of reference used in testing the null hypothesis, the parallel-group trial may be described as a placebo control, a non-treatment control, a historical control or an active control trial. In this article, we will examine the statistical properties of placebo-control, historical-control and active-control clinical trials for the testing of the null hypothesis. For simplicity and without loss of generality, we will focus on trials with no more than two different treatment arms.

## I. Placebo-Controlled Clinical Trial

For treating diseases or symptoms with a non-negligible placebo effect, placebo is often selected as the reference treatment. Let P denote the true response value of the placebo group, then the hypotheses are

$$\begin{aligned} H_0: T = P \text{ versus} \\ H_A: T > P \end{aligned} \quad (1).$$

The single most important statistical issue in designing a clinical trial is to minimize bias. Typically this is achieved by using the

blinding and randomization techniques. Such trials are often referred as well-controlled clinical trials.

In general, patients recruited into a clinical trial do not represent a random sample of the general patient population. Hence the generalization of the efficacy findings of the trial to general patient population should be considered with the medical feasibility, not purely through statistical justification. In addition, the effect size of the test treatment in the general population may not be covered by the estimation.

## II. Historical-Controlled Clinical Trial

Often there are situations in which placebo control becomes unsuitable as the reference treatment in a clinical trial because of ethical reason or patient recruitment difficulty. Alternative controls will be used instead. With historical control, patients are recruited to receive the test treatment in either a single treatment trial or to be randomized to receive one of the multiple formulations of the test treatments. In either setting, test treatment is compared with assumed placebo effect or untreated groups of the historical database. The comparison may be carried out by direct comparison across trials for testing

$$\begin{aligned} H_0: T = P_H \text{ versus} \\ H_A: T > P_H \end{aligned} \quad (2),$$

where  $P_H$  is the expected placebo effect of the historical trials, or by using a one-sample test against the following null hypothesis

$$H_0: T = P_0 \text{ versus}$$

$$H_A: T > P_0 \quad (3),$$

where  $P_0$  is the assumed expected value of placebo effect derived from historical data.

For testing against the null hypothesis (2), the asymptotic test uses the following test statistic

$$z = (\hat{T} - \hat{P}_H) / \text{s.e.}(\hat{T} - \hat{P}_H).$$

$H_0$  is rejected if  $z > z_{1-\alpha}$ .

For testing against  $H_0$  (3), the asymptotic test uses the following test statistic

$$z = (\hat{T} - P_0) / \text{s.e.}(\hat{T}).$$

Note that there are a few important differences among the hypotheses (1), (2) and (3). In testing hypothesis (1), the test treatment is compared with the concurrent placebo treatment  $P$  in the placebo control trial. In historical control trial, the test treatment is compared with the historical placebo. This difference plays an important role in the creditability of well-planned historical-controlled trials and the interpretation of the results. Let us examine the parameters to be tested in (1) and (2).

$$T - P = (T - P_H) + (P_H - P) \quad (4).$$

The term on the left hand side of (4) represents the parameter in (1) and the first term on the right hand side represents the parameter to be tested in (2),  $P$  can also represent the common expected non-existed placebo in historical trials. Hence, there is a bias of  $(P_H - P)$ , when using the comparison of test treatment with historical control to interpret the comparison in (2). The bias is expected to be small if historical trials are many and the confidence intervals of  $P_H$  are consistent across the historical trials. Often in this case, the bias is considered to be negligible. This is often called constancy

assumption of placebo mean. Under the constancy assumption,  $T - P$  can be estimated by  $\hat{T} - \hat{P}_H$  with  $\text{s.e.}(\hat{T} - \hat{P}_H) = \sqrt{[\text{Var}(\hat{T}) + \text{Var}(\hat{P}_H)]}$ .

The relationship of the parameters used in hypotheses (1) and (3) can be represented as follows,

$$T - P = (T - P_0) + (P_0 - P).$$

Hence the bias of using historical control trial to infer the difference between the test treatment and concurrent placebo is  $(P_0 - P)$ . Under the constancy assumption, the bias is 0.

The choice of  $P_0$  is crucial in this setting. In general, the variance of the estimate of the right hand side is  $\text{Var}(\hat{T}) + \text{Var}(\hat{P})$ . When  $P_0$  is the known true placebo mean and under the constancy assumption, bias  $(P_0 - P) = 0$  and  $T - P$  is estimated by  $\hat{T} - P_0$  with variance  $\text{Var}(\hat{T})$ . However, in practice,  $P_0$  is estimated from the data of historical trials. If the historical control trials ( $k=1, \dots, K$ ) share the same placebo mean, i.e.,  $P_0 = E(P_{HK})$ ,  $\text{Var}(\hat{P}_0) = [\sum_k n_k \text{Var}(\hat{P}_{HK})] / (\sum_k n_k)$ . However, if  $P_0 = P$ , but  $E(P_{HK}) \neq E(P_{HK'})$  for some  $k$  and  $k'$ ,  $\text{Var}(\hat{P}_0)$  can be estimated from a random effect model. In either case, a properly selected upper confidence limit  $\hat{P}_0^U$  of  $P_0$  is used. In the case when the number of historical trials is small and random effect model is not applicable, often the highest upper confidence limit of  $P_0$  among all historical trials is used.

### III. Active-Controlled Clinical Trial

An active-controlled clinical trial is often used when a placebo-controlled is not suitable while a standard treatment is available. The efficacy of the test treatment may be established by rejecting

$$H_0: T = A \text{ by showing } H_A: T > A,$$

which is essentially rejecting the null hypothesis in (1) with  $R=A$ , a standard active treatment. However, in many situations, the data do not support the superiority claim and the efficacy of the test treatment needs to be established through comparison with a non-concurrent placebo “P”,

$$\begin{aligned} H_0: T = \text{“P”} \text{ versus} \\ H_A: T > \text{“P”} \end{aligned} \quad (5).$$

This approach is often called non-inferiority testing. In order to infer the relationship between T and “P”, the active control A is served as a linkage in such a way

$$T - \text{“P”} = (T - A) + (A - A_H) + (A_H - P_H) + (P_H - \text{“P”}) \quad (6),$$

where A is the concurrent active control, P is the non-concurrent placebo,  $A_H$  is the historical active control and  $P_H$  is the historical placebo respectively.

Under the constancy assumption of the effect of active control,  $A - \text{“P”} = A_H - P_H$ , (6) can be rewritten as

$$T - \text{“P”} = (T - A) + (A_H - P_H) \quad (7).$$

Hence the hypotheses to be tested for efficacy are replaced by

$$\begin{aligned} H_0: T - A = - (A_H - P_H) \text{ versus} \\ H_A: T - A > - (A_H - P_H). \end{aligned}$$

A typical approach for inferring hypotheses (5) is to test the following hypotheses

$$\begin{aligned} H_0: T - A = - \delta \text{ versus} \\ H_A: T - A > - \delta \end{aligned} \quad (8),$$

where  $\delta$  is often called the non-inferiority margin and is determined to represent the value determined by  $(A_H - P_H)$ .

Note that:

Formulation in (7) is similar to (2) in that the parameters on the left hand side of the

hypotheses are of the current trial and the parameters on the right hand side of the hypotheses are of the historical control trials. The statistical considerations involved in testing the null hypothesis in (2) are also applicable here.

- Formulation in (8) is similar to (3) of historical control trials. Hence, the statistical considerations involving testing (3) are also applicable here.
- Bias, due to imbalance of baseline factors in historical control trials because of lack of randomization, may be minimized if not eliminated because of the randomization within the trials. However, the possible bias in comparison of the parameters of different clinical trials may still exist.
- In contrast to the constancy assumption on mean response of the treatment (i.e. placebo) in historical control trials, the constancy assumption is on the effect size of the active control.

Hence, when the number of historical control trials with placebo arm is large but the number of historical trials with the standard active treatment is small or the effect size of active treatment varies greatly among the historical trials, the advantage of active control design over historical control design may need to be reconsidered.

### Summary

We showed that strong assumptions are required beyond the simple blinding and randomization techniques to linking the current trial to historical data in order to be able to assess the efficacy of a test treatment in historical control trials or in non-inferiority active control trials. These assumptions are difficult to verify and hence become major weakness in interpreting the study results. The similarity with respect to the statistical issues between historical control design and non-inferiority active control design makes it

almost impossible to consider non-inferiority active control design differently from historical control design in both design and analysis.

\* The views expressed in the article are those of the authors and not necessarily of FDA. The work is partially supported by an FDA funded Review Science Research Grant (#RSR-01-20).

## **Active-control Trials: A Linguistic Problem**

By Janet Wittes, Ph.D.  
Statistics Collaborative, Inc.

Active-control Trials: Is the Problem Statistical?

When clinicians toss us the ball saying, “This is a statistical problem” and we statisticians lob it back with, “No, this is a clinical problem,” we are all in trouble. So it is with active controlled trials. The conundrum we face in designing and analyzing such trials stems, I believe, from tangled language and purpose. In these brief remarks, I will describe the linguistic problems we, both statisticians and clinicians, face in designing, analyzing, and interpreting an active controlled trial. Such trials come in two flavors, those that attempt to show superiority of the new intervention and those that aim to show “equivalence” or “non-inferiority”. The former poses little problem; the latter often leads us into an inferential morass.

To start with the easy case, suppose one is trying to demonstrate the superiority of a test drug over a standard treatment. The task is inferentially clear if previous randomized clinical trials have shown the standard treatment superior to placebo or even in the absence of data on its efficacy, if the standard is so widely used that control against placebo

would raise ethical eyebrows. For example, the CONVINCe trial, which is comparing controlled-onset extended-release (COER) verapamil to standard of care in persons with hypertension, is testing the hypothesis that COER-verapamil leads to a reduction in mortality relative to standard of care [1]. The trial is testing the simple null hypothesis that the incidence of fatal or nonfatal myocardial infarction, fatal or nonfatal stroke, or cardiovascular disease-related death in the two arms is equal against the natural two-sided alternative. If the data lead to rejection of the null hypothesis, then COER-verapamil will have been shown to lower the rate of clinical endpoints relative to the active control arm where, in this case, is standard of care. The fact that the trial has used an active control will have rendered the conclusion more relevant to practice than had the trial used a placebo.

The situation becomes much more difficult for trials that aim to show “equivalence” or “non-inferiority” of the test agent to the active comparator. In discussing such trials, we must come to grips with the deceptive language inherent in the terms “equivalence” and “non-inferiority”. A normal person might well assume that “equivalent” means “the same as” and “non-inferior” means “no worse than”; in clinical trial-speak, however, “equivalent” means “not unacceptably different from” and “non-inferior” means “not unacceptably worse than”. Our failure to adopt honest language obfuscates the hypotheses we are testing and casts an unrealistically favorable light upon our results. The reader might say, “yes”, but deceptive language lies at the root of misinterpretation.

The “not unacceptably worse than” trial is typically relevant to clinical endpoints while the “not unacceptably different” trial applies to surrogate markers. Consider a trial of a new

vaccine for influenza compared to one already marketed. If the endpoint is incidence of disease, then the appropriate design is the “not unacceptably worse than” trial. Suppose the marketed product has an efficacy of 70 percent where vaccine efficacy is defined as 1-incidence in vaccine/incidence in placebo. If recipients are expected to find the new vaccine more tolerable, public health workers may believe that an efficacy of only 60 percent is “not unacceptably worse than” than the marketed vaccine and therefore may urge approval of such a vaccine. Of course, if the new vaccine shows higher efficacy than the marketed one, so much the better. The “not too much worse than” trial allows rejection of the null hypothesis if the new vaccine is in fact superior to the previous one.

On the other hand, if the endpoint is not clinical efficacy but some marker of immune response, then one might demand that the new vaccine have properties similar to that of the marketed agent. One would specify a range that represents “not unacceptably different from”. For example, the marketed agent may confer a mean eight-fold increase in an immune marker; the “not unacceptably different” interval may be four-fold to sixteen-fold. Observing a 32-fold increase in the marker would then be a failure to show “not unacceptably different from” even if, at first blush, a 32-fold increase seems intuitively better than an eight-fold increase. The problem with presuming larger means better in the absence of data is that one cannot be certain how changes in surrogate markers affect changes in clinical endpoints.

Many people have dealt with the challenges in designing a “not unacceptably different from trial” (a.k.a., an equivalence trial) abound. One must select the margin of indifference narrow enough to exclude a clinically relevant degradation of effect but wide enough to allow a feasible sample size [2-4]. As Temple and Ellenberg have described, one must think

about “assay variability”, the possibility that the comparator agent is less effective in the trial being designed than in the trials that demonstrated its efficacy [5]. One must avoid finding no difference from standard by virtue of running a sloppy trial; in a superiority trial, sloppiness pulls the data towards the null hypothesis while in a “not unacceptably worse” trial, sloppiness makes two otherwise different drugs appear more similar than they in fact are.

Finally, both clinicians and statisticians should question the purpose of the trial. If the margin of “equivalence” is too wide and the active control too variable in its efficacy, the data may formally lead to rejection of the null hypothesis of “equivalence” but fail to show convincingly that the test agent is an acceptable substitute for the agent already in use. We as statisticians must not allow our well-honed statistical machinery to appear to “prove” a hypothesis that is not clinically important. And let us resist the words “equivalence” and “non-inferiority”, for they mean not what they say.

## REFERENCES

1. Black H, Elliott W, *et al.* (1998). Rationale and design for the Controlled ONset Verapamil INvestigation of Cardiovascular Endpoints (CONVINCE) Trial. *Control Clin Trials*, 19:370-390.
2. Blackwelder W (1982). "Proving the null hypothesis" in clinical trials. *Control Clin Trials*, 3:345-353.
3. Schuirmann D (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J Pharmacokinet Biopharm*, 15:657-680.
4. Makuch R, Simon R (1978). Sample size requirements for evaluating a conservative therapy. *Cancer Treat Rep*, 62:1037-40.
5. Temple R, Ellenberg S (2000). Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues. *Ann Int Med*, 133:455-63.

□□□□□

# MESSAGE FROM THE PRESIDENT

January 2001

Dear ICSA Members:

The ICSA is moving into the twenty-first century. Undoubtedly, this is an era of information. We are facing huge amount of data, generated from almost every field of sciences. How to extract useful information and transfer into knowledge will be the responsibility of the 21<sup>st</sup> century statisticians. We should feel lucky as a part of the exciting discipline.

The ICSA has done an excellent job in the past decade. Last year the ICSA has made some big progress under the leadership of Professor Chien-Pai Han, especially in building bridges between ICSA and the other statistics community. Through the efforts of Chien-Pai, Tim Chen and many of our members, we obtained the encouragement from the president of ASA to nominate ICSA members as candidates for ASA president-elect. This year we will continue to interact with other statistical societies, not only in the North America, but also in the pacific region.

Bridging will be the key word in developing many of our programs. For example, the editor of ICSA Bulletin, Dr. Sue-Jane Wang, has done a great job to create a new style, which I believe will bridge the ICSA members from different regions. This year we will make effort to enhance our Web site. Internet is no doubt the most efficient way in global communication. Any suggestion to enhance the content of ICSA web pages is highly appreciated.

Our discipline has shown a big impact to the industry in the North America. ICSA members should help other regions to establish linkage among industry, government and academics. This can be gradually achieved by establishing collaboration among statisticians from different regions. Workshops and conferences across the regions could be a starting point. In the workshops or conferences, a forum to discuss new developments in statistical science can be formed and valuable interdisciplinary activities can be organized.

A good infrastructure will be essential for ICSA to become an efficient organization to all members from different regions and disciplines. I will encourage members to submit their proposals to establish new chapters or sections. With chapters established in different regions, ICSA will become a truly international statistical society.

To enhance the skills of ICSA members in career relevant areas, continuing education will be maintained and enhanced through short courses and training programs offered across the regions. The short courses organized by the Applied Statistics Symposium Committee have been recognized to be a successful program. This kind of program will also be activated in other regions besides in North America.

As for our publications, according to the Editor-in-chief of Statistica Sinica (SS), Professors Ker-Chau Li and Yi-Ching Yao, the submission to SS is growing steadily. More high quality research papers, no matter in methodology or application, are welcome submitted to SS.

As you know, the term of our Executive Director (ED), Dr. Naitee Ting, ends last December. Our new Executive Director, Dr. Yi Tsong, has been elected. Dr. Ting has contributed a lot to ICSA. He and Dr. Tsong have made a very smooth transition. Dr. Tsong has always been very active. We are lucky to have him to be our new ED. If you have any suggestions about ICSA's activities, you can either e-mail me ([hsiang1@nhri.org.tw](mailto:hsiang1@nhri.org.tw)) or Dr. Tsong ([tsong@cder.fda.gov](mailto:tsong@cder.fda.gov)).

Best wishes for a very productive new year,

*Chao Agnes Hsiung*     **President**

## Special Thanks from The Editorial Board

The Editorial Board would like to thank Dr. Chien-Pai Han, our past President, for his timely support during the preparation of this issue. We are also indebted to Dr. Timothy Chen, our President of 1999 for his enthusiastic and timely contributions.

If you have a new idea and are interested in joining us, please send your C.V. including your plan to the Editorial Board [WANGS@CDER.FDA.GOV](mailto:WANGS@CDER.FDA.GOV) for consideration.

We encourage your active involvement in the ICSA Bulletin. Every effort counts.



□□□□□