STATISTICAL CONSIDERATIONS ON PHARMACODYNAMIC ASSESSMENT OF HUMAN ABUSE POTENTIAL STUDIES

Ling Chen, Ph.D.
FDA/CDER/OTS/DBVI

CCALC/CSS, 10/11/2018

Disclaimer

This presentation reflects the views of the author and should not be construed to represent FDA's views or policies.

Outline

- Margins in statistical testing
- Approaches to handle secondary endpoints
- Statistical analyses
  - Sampling distributions of t-type random variable
  - Nonparametric tests
- Remarks

Hypotheses

- FDA 2017 Guidance recommends hierarchically testing the following hypotheses:

1. $H_0 : \mu_C - \mu_P \le \delta_1$ versus $H_a : \mu_C - \mu_P > \delta_1$ where $\delta_1 > 0$.
2. $H_0 : \mu_C - \mu_T \le \delta_2$ versus $H_a : \mu_C - \mu_T > \delta_2$ where $\delta_2 \ge 0$.
3. $H_0 : \mu_T - \mu_P \ge \delta_3$ versus $H_a : \mu_T - \mu_P < \delta_3$ where $\delta_3 > 0$.

- The actual values of δs vary according to such factors as subjective measures, drug class, and route of drug administration. All the margins should be pre-specified and justified in the protocol. The nominal type I error rate for each test is 0.05.

How to define δs

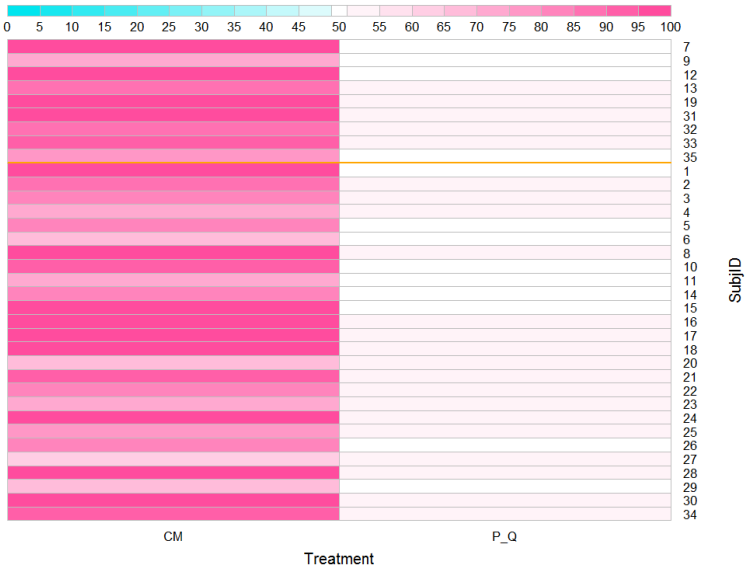
- Each δ represents a clinically meaningful difference between two treatments in a comparison.
- For example:
  - Whether a subject responds to a positive control should be primarily evaluated using a bipolar Drug Liking VAS of 0-100, such that placebo should produce a score between 40-60 points, the positive control should produce a score outside of the placebo range; and there should be a difference of at least 15 points between placebo and the positive control response. Should 15 also be used for the margin of the validation test in the Treatment Phase? (δ_1)
  - How large difference in means of maximum liking, between a test drug and alprazolam, should be so that a test drug will be considered having less abuse potential than alprazolam? (δ_2)
- These questions essentially are not for statisticians.

About δ_1

- I believe that at least 15 points difference, between the positive control and placebo used in the Qualification Phase for Drug Liking Emax, was originally proposed by pharmacologists and/or clinicians.
- Logically, because all subjects randomized to the Treatment Phase passed the Qualification Phase, **most subjects** in the Treatment Phase should have a maximum liking to the positive control at least 15 points larger compared to placebo. Then, it should not be a problem of using $\delta_1 = 15$ for the validation test in the primary analysis.
- However, ….

Heat map display by treatment

Drug Liking Emax (Qualification Phase)



Drug Liking Emax (Treatment Phase)

Continued…

- The problem of failing the validation test may not be or may not only be due to the margin selection but due to the design of the Qualification Phase.
- Should both doses or at least one dose of positive control used in the Treatment Phase be used in the Qualification Phase?
- Should the Qualification Phase not use 2x2 design? Because of the 2x2 design, the probability that a subject through guessing gives proper answers to both positive control and placebo is 0.5, as subjects are informed what treatments will be used in the Qualification Phase.

Let A and B be events that a subject gives proper answer to the first and second treatments, respectively. Because the subject knows what treatments will be used in the qualification phase,

$$P(A \cap B) = P(A)P(B|A) = 0.5 \times 1 = 0.5$$

Secondary Analysis

- The sample size is determined based on the comparisons for the primary endpoint(s). Therefore, when using δs similar to those in the primary analysis, the secondary analysis is often under powered.
- I suggest prespecifying two or three key secondary endpoints in the study. The key secondary endpoints should be directly related to the specific pharmacological effects of the test drug.
- If δs are set as zero in the secondary analysis, the lower 95% confidence interval (CI) limit for upper-tail test, and the upper 95% CI limit for lower-tail test, may provide some information about δs.
- When a two-sided test with margin zero is used for the comparison between test drug and placebo, the nominal type I error rate should be 0.10, because in such a case, the research hypothesis is the null hypothesis.
- We may discuss whether it would be sufficient to provide descriptive statistics for other endpoints (secondary and exploratory).

Regarding the Robustness of Mixed Effects Model

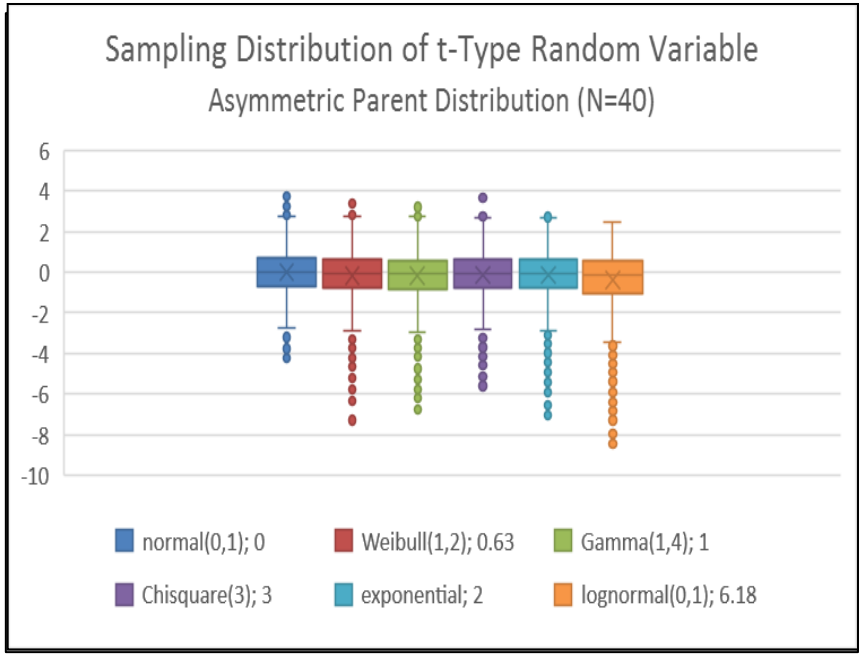
- People may be interested in the robustness of the mixed effects model, if the error term is not normally distributed, and sample sizes are around 30-40.
- This is one of the ongoing research projects conducted by the Agency. Before the research is completed and thoroughly evaluated, and conclude the robustness of the mixed effects model, the normality assumption should be checked before using the mixed effects model.

Testing for mean difference

- If the normality assumption of the mixed effects model is violated, the first choice for testing the difference in means between two treatments is the paired-t test.
- The t- test is robust, when the parent distribution is symmetric.
- However, …

Sampling Distribution of t-Type Random Variable

- The t-type random variable: $t_{type} = \frac{\sqrt{N}(\bar{X}-\mu)}{S}$
- The skewness of the parent distribution has a greater effect on the distribution of t_{type} than the kurtosis does, and the positive skewness in the parent distribution results in the sampling distribution of t_{type} being negatively skewed. (See Neyman and Pearson (1928) and Pearson (1928,1929)).
- When the parent distribution is positively skewed, the short right tail of the sampling distribution of t_{type} leads to a loss of power for the upper-tail test of the population mean. The long left tail of the sampling distribution of t_{type} leads to an inflated type I error rate for the lower-tail test of the population mean.



Replication=10000. The legend shows the corresponding parent distributions. For skewed parent distributions, the skewness of each distribution is also listed next to the parent distribution.

About t test

- In the primary analysis, except for the comparison between test drug and placebo, the tests for the other comparisons are upper-tail tests.
- Johnson (1978), Sutton (1993) and Chen (1995) studied upper-tail t test, and proposed modified t tests for the mean of positively skewed distributions. Comparisons of Johnson's t test, Sutton's composite test and Chen's t_2 test can be found in Chen (1995).
- Zhou and Gao (2000) studied one-sided confidence intervals for the mean of positively skewed distributions, and recommended the use of the bootstrap version of Hall's(1992) transformation approach for construction of one-sided confidence interval when data follow a positively skewed distribution.
- If the distribution of paired difference is negatively skewed and the test is an upper-tail test, or if the distribution of paired difference is positively skewed and the test is a lower-tail test, the type I error rate of the t test is inflated (See Sutton, 1993).

Nonparametric Tests

1. When t test cannot be used, the Sign test is often used for testing the median of difference between two treatments. The Sign test is a very simple test.
   - When the Sign test is used, the sample median of paired differences is not of interest.
   - The calculation of the Sign test and the confidence interval for median based on the Sign test should exclude subjects who had zero difference in Emax scores between two treatments. (Daniel, 1990)
2. The bootstrap test
   – Sutton (1993) studied bootstrap methods for testing the mean of positively skewed distribution.
   – Sutton (1993) reported that “None of the bootstrap procedures examined can be deemed highly accurate for lower-tail tests in all of the situations considered; …”
3. The permutation test for paired samples
   – It is computational intensive.
   – It has pros and cons (Berger, 2000)

Remarks

- The current design for Qualification Phase is questionable.
- The t test is robust, when the parent distribution is symmetric. Therefore, the Wilcoxon Signed Rank test should not be used, when the distribution of paired difference is symmetric.
- If the normality assumption of the model is violated, there are several methods (parametric or nonparametric) which may be used to test treatment differences. However, any test which will inflate type I error rate is not allowed.
- Both mean and median are measures of the central tendency of a distribution. If the median is of interest in HAP studies, when the distribution of paired difference is skewed (regardless positively skewed or negatively skewed), nonparametric tests for testing median are preferred (regardless upper-tail, lower-tail or two-tail test).
- We will consider and evaluate any statistical method that you may propose.

Reference

1. FDA Guidance for Industry: Assessment of Abuse Potential for Drugs, January 2018.
2. Neyman, J., and Pearson, E.S. (1928). "On the Use of Interpretation of Certain Test Criteria for Purposes of Statistical Inference, Part I," *Biometrika, 20A,* 175-240.
3. Person, E. W. (1929), "The Distribution of Frequency Constants in Small Samples from Non-Normal Symmetrical and Skew Populations," *Biometrika,* 21, 259-286.
4. Johnson, N. J. (1978). "*Modified t Tests and Confidence Intervals for Asymmetrical Populations*," *Journal of the American Statistical Association*, Vol. 73, No.363, 536-544.
5. Sutton, C. D. (1993). " Computer-Intensive Methods for Tests About the Mean of an Asymmetrical Distribution," *Journal of the American Statistical Association*, Vol. 88, No. 423, 802-810.
6. Chen, L. (1995). "Testing the Mean of Skewed Distributions," *Journal of the American Statistical Association*, Vol. 90, No. 430, 767-772.
7. Zhou, X. and Gao, S. (2000). *The American Statistician*, Vol. 54, No.2, 100-104.
8. Hall, P. (1992). "On the Removal of Skewness by Transformation," *Journal of Royal Statistical Society*, Series B, 54, pp221-228.
9. Daniel, W. W. (1990). *Applied Nonparametric Statistics, Second Edition*, PWS-KENT Publishing Company.
10. Berger, V. W. (2000). "Pros and Cons of Permutation Tests in Clinical Trials," *Statistics in Medicine*, 19: 1319-1328.

Thank you!