THE IMPACT OF MODEL ASSUMPTIONS IN HUMAN ABUSE POTENTIAL STUDIES

Ran Bi & Qianyu Dang
FDA/CDER/OTS/OB/DBVI

Disclaimer

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

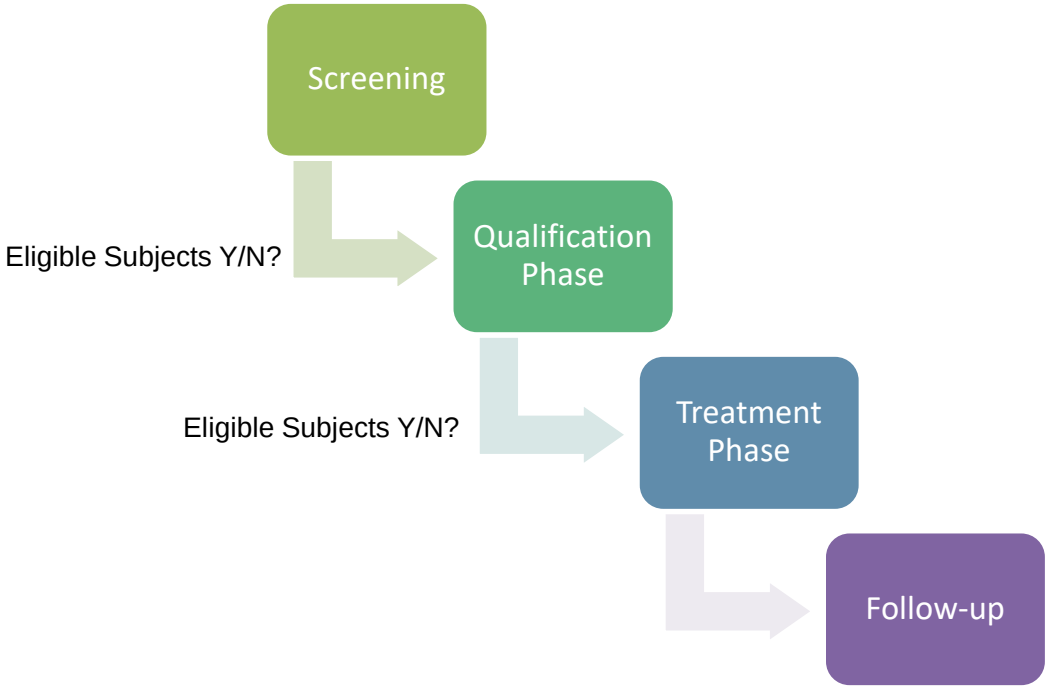
Outline

- Introduction to HAP Studies
- Statistical Modeling Procedure
- Key Questions
- Simulations
- Discussion

Human Abuse Potential (HAP) Study

- A human abuse potential (HAP) study is a type of early phase clinical study.
- It evaluates whether or the magnitude of a drug produces subjective abuse-related responses through the drug effects on the human central nerve system (CNS).
- It assesses abuse potential in individuals with a history of recreational use of drugs of abuse.
- Important tools to assess the subjective drug effects include Visual Analogue Scales (VAS) of drug liking, take drug again, good/bad drug effects, etc.

Design of a HAP study



Design of a HAP study

- Treatment phase usually include placebo (P), positive control (C) and different doses of test drug (T).
- Typically it is a randomized, double-blinded crossover study with balanced William Square Design.
- Sample size is usually 30 ~ 50.

Hypotheses Testing (FDA 2017 Guidance)

1. Validation test of the sensitivity and integrity of the study: Does the positive control (C) produce mean responses that show greater abuse potential compared to placebo (P)? Thus, the hypothesis should be tested as following:

$$H_0\colon \mu_C - \mu_P \leq \delta_1 \quad \text{versus} \quad H_a\colon \mu_C - \mu_P > \delta_1 \quad \text{where} \quad \delta_1 > 0$$

   Typically $\delta_1 = 15$ for Drug Liking Emax.

2. Does the test drug (T) produce mean responses that show less abuse potential compared to positive control?

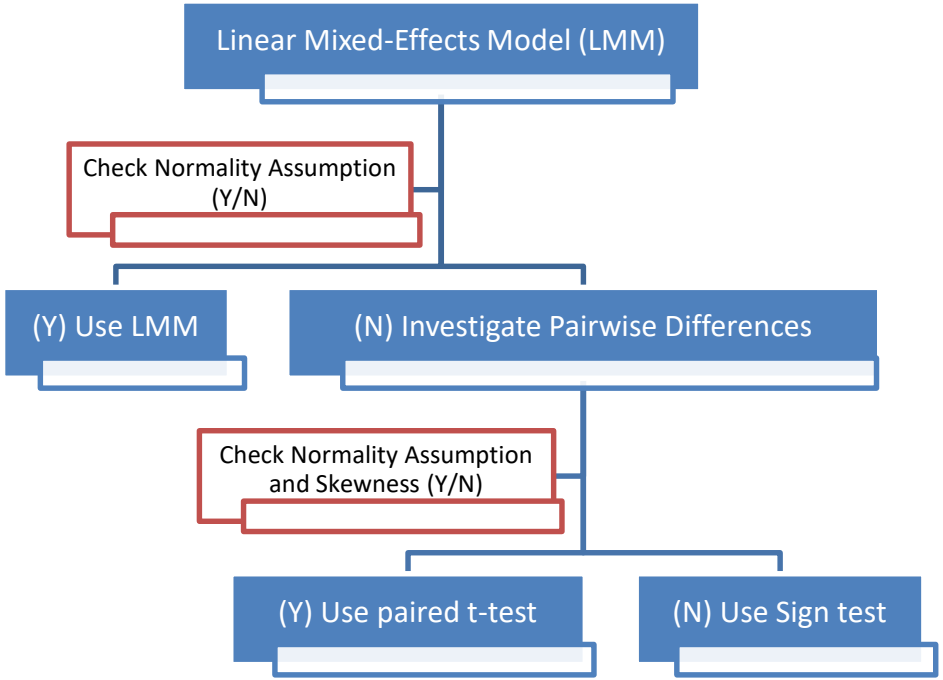
$$H_0\colon \mu_C - \mu_T \leq \delta_2 \quad \text{versus} \quad H_a\colon \mu_C - \mu_T > \delta_2 \quad \text{where} \quad \delta_2 \geq 0$$

3. Does the test drug produce mean responses that show similar abuse potential compared to placebo?

$$H_0\colon \mu_T - \mu_P \geq \delta_3 \quad \text{versus} \quad H_a\colon \mu_T - \mu_P < \delta_3 \quad \text{where} \quad \delta_3 > 0$$

   Usually $\delta_3 = 11$ for Drug Liking Emax.

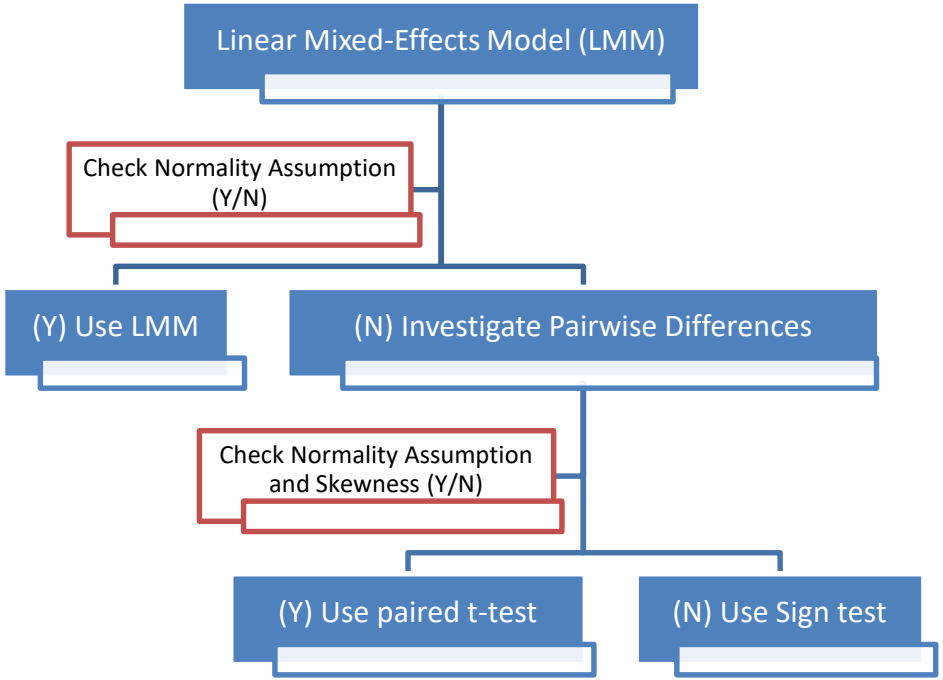
Statistical Modeling in HAP Studies



Statistical Analyses in HAP Studies

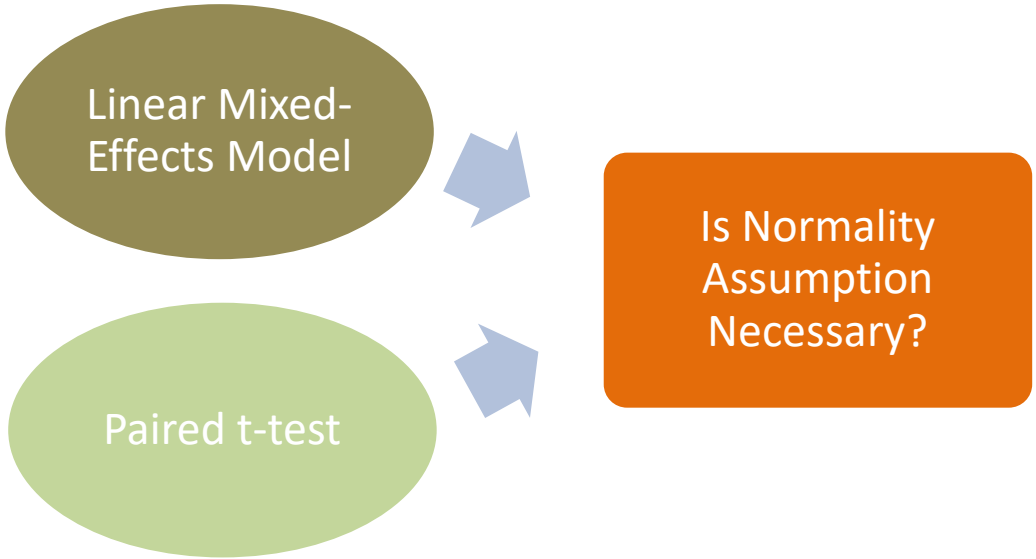
- The statistical model that should be used in HAP studies is a linear mixed-effects model, which includes period, sequence, treatment, and first-order carryover effect as fixed effects, and subject as a random effect.
- Assumptions for linear mixed model:
  - Linear relationship between response and explanatory variables.
  - Homogeneity of variance.
  - The model errors (aka conditional residuals) are independent and normally distributed.
  - The random effects are normally distributed.
  - The random effects and model errors should also be independent of each other.
- The normality assumption should be investigated using Shapiro-Wilk W-test.
- The impact of error term distribution is our main focus in this research

Statistical Modeling in HAP Studies



Key Questions

Is the current practice the best approach?



Linear Mixed-Effects Model in HAP Study

In matrix formulation,

$$Y = X\beta + Zu + \epsilon$$

Y : $(3m \times 1)$ vector of response variables

X: $(3m \times 3)$ design matrix for fixed effects

β: (3×1) vector of fixed effects (treatments), indicates Positive Control, Test Drug, and Placebo respectively

Z: $(3m \times m)$ design matrix for random effects

u: $(m \times 1)$ vector of random effect (subjects), $u \sim N_m(0, G)$, $G = \sigma_u^2 I_m$

ϵ: $(3m \times 1)$ vector of within-subject error term, $\epsilon \sim N_{3m}(0, R)$, $R = I_m \otimes R_i$

$$R_i = \begin{pmatrix} \sigma_{e_1}^2 & 0 & 0 \\ 0 & \sigma_{e_2}^2 & 0 \\ 0 & 0 & \sigma_{e_3}^2 \end{pmatrix}$$

Impact of Normal Assumption

- Weighted least square estimates (weighted LSE) for fixed effects:

$$\hat{\beta} = \left(X^T \hat{V}^{-1} X\right)^{-1} X^T \hat{V}^{-1} Y,$$

which is identical to the maximum likelihood estimates (MLE), $V = ZGZ^T + R$.

The unbiasedness of $\widehat{\boldsymbol{\beta}}$ requires only the assumption of $\boldsymbol{E(Y) = X\beta}$, even if the assumptions about random effects and residuals are inappropriate.

- The variance of $\hat{\beta}$ is estimated by:

$$Var\left(\hat{\beta}\right) = \left(X^T \hat{V}^{-1} X\right)^{-1}.$$

The MLE estimation of $\widehat{\boldsymbol{V}}$ has no closed form and has to be estimated iteratively, its convergence relies on correct specification of the covariance structure and the normality assumption is needed for its correct estimation. Hence, normality assumptions are needed for inference on the test of fixed effects.

Impact of Normal Assumption

$$Y = X\beta + Zu + \epsilon$$

- The unbiasedness of $\hat{\beta}$ is not affected by the normality of error term ϵ.
- The estimate of $Var(\hat{\beta})$ is affected by distribution of ϵ and the covariance structure. Deviation from normal may affect the inference.
- In a HAP study, the hypothesis testing is for the mean difference (estimated by LS mean) between treatment groups based on t statistics: $\frac{Means(trt1)-Means(trt2)}{SE[Means(trt1)-Means(trt2)]}$. The numerator is estimated from $\hat{\beta}$, while the denominator is related to $Var(\hat{\beta})$, thus the t statistic is affected by the normality of error term ϵ.
- The asymptotic property may not fully apply for such sample size.

Simulation Settings

- All the parameters used are based on typical HAP studies we reviewed
- Simulation of a hypothetical study: A 3-way crossover study with 36 subjects
  - ❖ Fixed effects (3 treatments):
    - o Positive Control (C): mean = 65
    - o Test Drug (T): mean = 61
    - o Placebo (P): mean = 50
  - ❖ Random effects (subjects): $u_i \sim N(0, 7)$
  - ❖ The model error ϵ was simulated from various distributions, and scaled to have mean 0 and standard deviations $\sigma_{e_1} = 15$, $\sigma_{e_2} = 10$, $\sigma_{e_3} = 5$ for Positive Control, Test Drug, and Placebo respectively

Simulation Settings

- The model error ϵ was simulated from various distributions, and scaled to have mean 0 and standard deviations $\sigma_{e_1} = 15$, $\sigma_{e_2} = 10$, $\sigma_{e_3} = 5$ for Positive Control, Test Drug, and Placebo respectively

  - Normal
  - Log-Normal with skewness from 0.4 to 8
  - - Log-Normal with skewness = -8
  - t_5
  - Exponential(1), skewness around 2
  - χ^2_3, skewness = 1.63;
    χ^2_5, skewness = 1.26;
    χ^2_{10}, skewness = 0.89
  - Weibull(1, 0.6), skewness = 4.6;
    Weibull(1, 1), skewness = 2;
    Weibull(1, 1.5), skewness = 1.07;
    Weibull(1, 3.6), skewness = 0;
    Weibull(1, 10), skewness = -0.64;
    Weibull(1, 100), skewness = -1.08

  - Other distributions include bi-model and Laplace

Simulation Procedure

- For each simulated dataset, fit a linear mixed-effects model to the response with treatments as fixed effects and subjects as random effects.
- Calculate the t test statistic for each paired treatment difference, and determine whether the null hypothesis is rejected or not by comparing the test statistic with critical value:
  - C – P: the difference between positive control and placebo, upper-tail test with test value of 15
  - C – T: the difference between positive control and test drug, upper-tail test with test value of 4
  - T – P: the difference between test drug and placebo, lower-tail test with test value of 11
- Repeat the data simulation and hypotheses testing procedures for 10,000 times, and calculate the percentage of rejections among 10,000 simulations, indicating type I error rate.

Simulation Results

Linear mixed-effects model, $\alpha = 0.05, n = 36$

Error Distribution	Actual Skewness (overall residuals)	Actual Skewness (C-P difference of residuals)	Type I error rate (C-P)	Actual Skewness (T-P difference of residuals)	Type I error rate (T-P)
Normal	-0.003	-0.002	0.050	0.005	0.051
t_5	-0.018	-0.023	0.053	-0.038	0.053
Weibull(1, 2.5)	0.429	0.288	0.046	0.230	0.054
Log-Normal	0.603	0.413	0.044	0.305	0.058
Log-Normal	0.942	0.658	0.040	0.483	0.061
Weibull(1, 1.5)	1.266	0.888	0.037	0.675	0.064
Chi-square(3)	1.959	1.330	0.034	1.017	0.068
Exp(1)	2.443	1.584	0.029	1.252	0.075
Weibull(1, 0.6)	5.653	3.616	0.016	2.659	0.105
Log-Normal	9.295	6.035	0.015	4.595	0.107
Weibull(1, 6)	-0.456	-0.298	0.057	-0.229	0.046
Weibull(1, 10)	-0.753	-0.508	0.061	-0.398	0.044
Weibull(1, 100)	-1.286	-0.901	0.069	-0.691	0.041
-Log-Normal	-8.203	-5.401	0.138	-4.184	0.021

Check distribution of overall residuals or difference of paired residuals? $\alpha = 0.05, n = 36$

Treatment	Error Distribution	Skewness	Actual Skewness (overall residuals)	Actual Skewness (C-P difference of residuals)	Type I error rate (C-P)
C	Weibull(1, 10)	-0.64	-0.043	-0.544	0.060
T	Weibull(1, 1)	2			
P	Normal	0			

Check distribution of overall residuals or difference of paired residuals? $\alpha = 0.05, n = 36$

Treatment	Error Distribution	Skewness	Actual Skewness (overall residuals)	Actual Skewness (C-P difference of residuals)	Type I error rate (C-P)
C	Weibull(1, 10)	-0.64	-0.043	-0.544	0.060
T	Weibull(1, 1)	2			
P	Normal	0			
C	Weibull(1, 2.5)	0.36	0.336	0.605	0.042
T	Log-Normal_0.8	0.8			
P	-Log-Normal_8	-8			
C	Weibull(1, 6)	-0.37	0.962	-0.372	0.057
T	Weibull(1, 0.6)	4.6			
P	Log-Normal_2	2			
C	Weibull(1, 10)	-1.08	-0.726	-0.936	0.068
T	Log-Normal_0.8	0.8			
P	Log-Normal_0.5	0.5			
C	Chi-square(3)	1.63	-2.777	1.392	0.035
T	-Log-Normal_8	-8			
P	t_5	0			

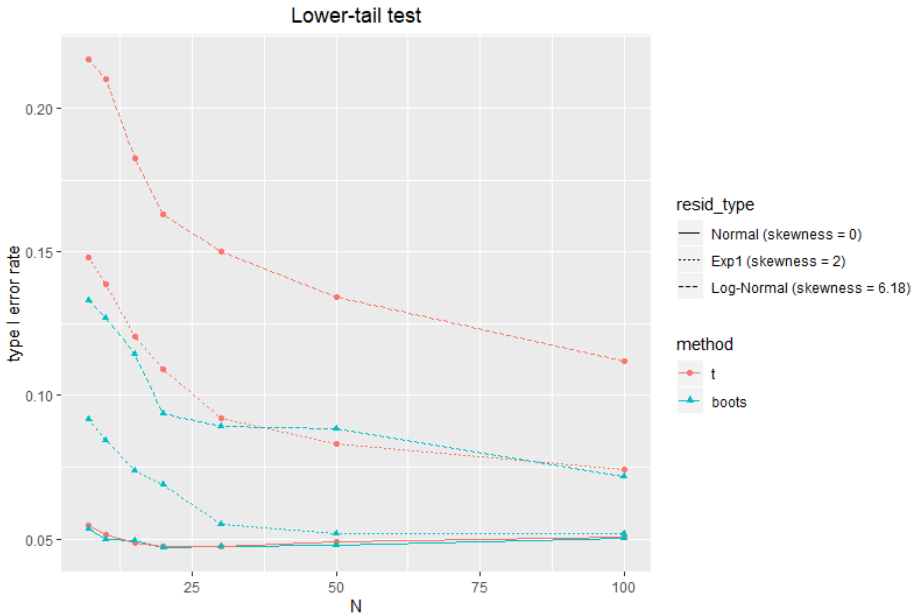
Simulation Results

Paired t-test, $\alpha = 0.05, n = 36$

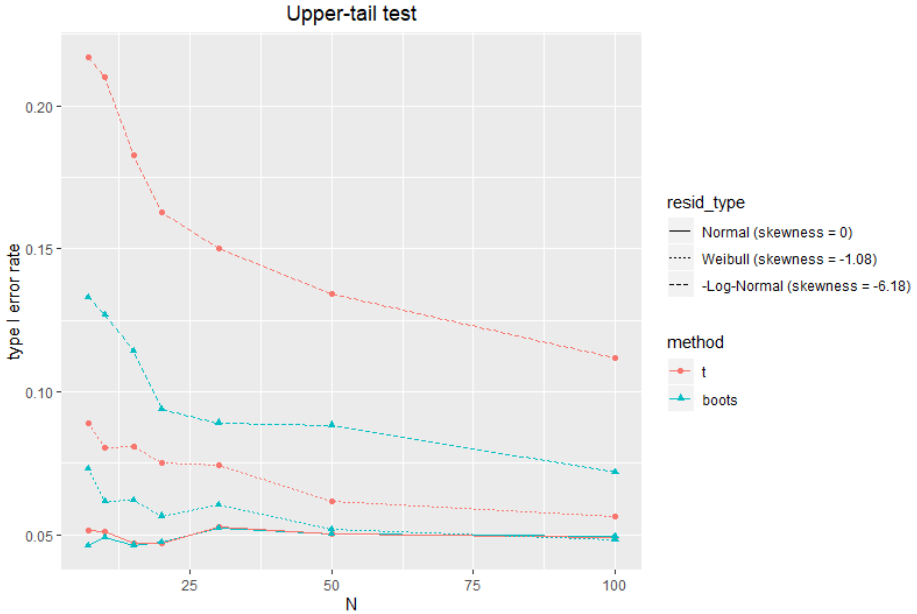
Error Distribution	Actual Skewness (C-P)	Type I error rate (C-P) (with Bootstrapping*)	Actual Skewness (T-P)	Type I error rate (T-P) (with Bootstrapping)
Normal	-0.002	0.044 (0.046)	0.005	0.051 (0.050)
t_5	-0.023	0.047 (0.054)	-0.038	0.048 (0.055)
Weibull(1, 2.5)	0.288	0.044 (0.049)	0.230	0.06 (0.054)
Log-Normal	0.413	0.041 (0.045)	0.305	0.057 (0.055)
Log-Normal	0.658	0.039 (0.049)	0.483	0.060 (0.055)
Weibull(1, 1.5)	0.888	0.040 (0.054)	0.675	0.069 (0.054)
Chi-square(3)	1.330	0.032 (0.043)	1.017	0.065 (0.057)
Exp(1)	1.584	0.031 (0.052)	1.252	0.071 (0.059)
Weibull(1, 0.6)	3.616	0.024 (0.057)	2.659	0.11 (0.104)
Log-Normal	6.035	0.019 (0.048)	4.595	0.099 (0.100)
Weibull(1, 6)	-0.298	0.057 (0.051)	-0.229	0.050 (0.053)
Weibull(1, 10)	-0.508	0.061 (0.054)	-0.398	0.050 (0.054)
Weibull(1, 100)	-0.901	0.064 (0.055)	-0.691	0.046 (0.052)
-Log-Normal	-5.401	0.138 (0.111)	-4.184	0.020 (0.054)

*10,000 Bootstrap samples were used to get an average type I error rate

Impact of Sample Size: Upper-tail and Lower-tail Test



Estimated Type I Error Rates vs. sample size for Both Paired t-test and Bootstrapping t-test with positive skewed paired difference in lower-tailed tests

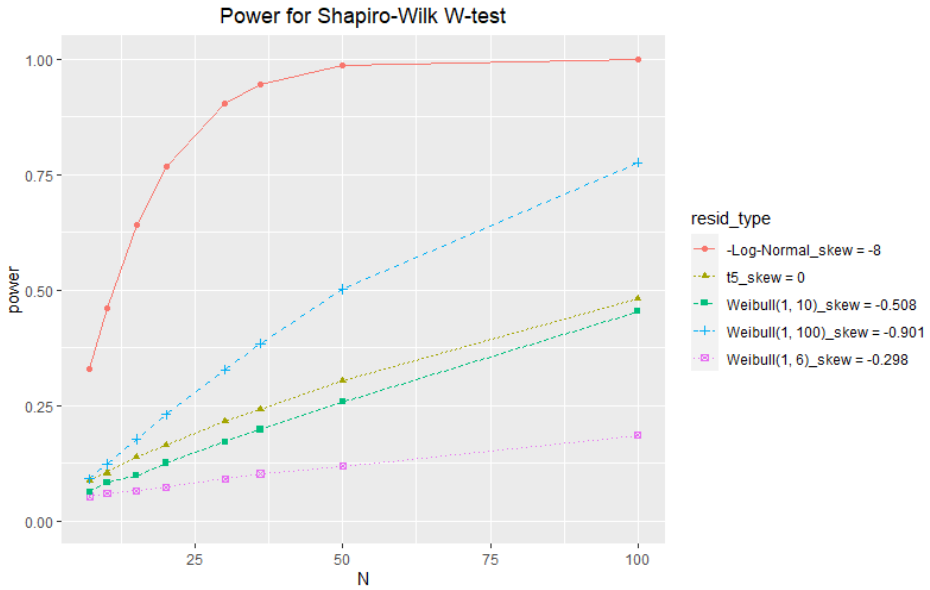


Estimated Type I Error Rates vs. sample size for Both Paired t-test and Bootstrapping t-test with negative skewed paired difference in upper-tailed tests

Remarks

- In the linear mixed-effects model, the test results may be more related to the distribution of the difference in paired residuals, while in current approach we just check the distribution of all residuals.
- The type I error rate can differ substantially from the nominal level if the skewness of paired difference is severe, even as the sample size to be as large as 100. As the distribution gets more skewed, the type I error rate gets further away from the nominal level.
- In paired t-test, bootstrapping can have the type I error rate closer to its nominal level when the skewness of paired difference is not severe.

Power for Shapiro-Wilk W test



- Razali and Wah (2011) stated that “Shapiro-Wilk W-test is the most powerful test for both symmetric non-normal and asymmetric distributions. However, the power of Shapiro-Wilk W-test is still low for small sample size.”

Conclusions

- The normality assumptions need to be checked before applying linear mixed-effects model and paired t-test.
- In the linear mixed-effects model approach with small sample size of 30 ~ 50, the distribution of the difference in paired residuals for each contrast may be more relevant.
- The type I error rate will be inflated in lower tailed-test for positive skewness and upper tailed-test for negative skewness.
- The type I error rate may be inflated for both mixed model and paired t-test with a sample size less than 40 even when the distribution is quite symmetric (skewness = -0.5 to 0.5), though the magnitude of inflation may not be large.
- When the paired difference has skewness between -1 and 1, Bootstrapping t-test performs better than paired t-test in reducing the inflation of type I error rate.
- If the distribution is quite symmetric, the normality likely will not be rejected by the W-test.

Acknowledgement

- Dr. Ling Chen, DB VI
- Dr. Anna Sun, DB VI
- Dr. Wei Liu, DB VI
- Dr. Mohammad Rahman, DB VI Deputy Director
- Dr. Yi Tsong, DB VI Director

References

1. FDA Guidance for Industry: Assessment of Abuse Potential for Drugs, January 2018.
2. Verbeke, G. and Lesaffre, E. 1997. "The effect of misspecifying the random effects distribution in linear mixed models for longitudinal data". *Comput. Statist. Data Anal.* 23, 541-556.
3. 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.
4. 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.
5. Chen, L. (1995). "Testing the Mean of Skewed Distributions". *Journal of the American Statistical Association*, Vol. 90, No. 430, 767-772.
6. Zhou, X. and Gao, S. (2000). "One-sided confidence intervals for means of positively skewed distributions". *The American Statistician*, Vol. 54, No.2, 100-104.
7. Hall, P. (1992). "On the Removal of Skewness by Transformation". *Journal of Royal Statistical Society*, Series B, 54, pp221-228.
8. Razali, N. M. and Wah, Y. B. (2011). "Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests". *Journal of Statistical Modeling and Analytics*, 2(1), 21-33.

Thank You!

Bootstrapping Method

1) Calculate the paired difference $x_1, x_2, \ldots, x_N$ between two treatments;
2) Draw N Bootstrap samples with replacement from $x_1, x_2, \ldots, x_N$, denote as $x_1^*, x_2^*, \ldots, x_N^*$;
3) Compute

$$t_b^* = \frac{\widehat{\theta_b^*} - \hat{\theta}}{\widehat{\sigma_b^*}},$$

where $\widehat{\theta_b^*}$ and $\widehat{\sigma_b^*}$ are the sample mean and standard error of Bootstrap samples $x_1^*, x_2^*, \ldots, x_N^*$, $\hat{\theta}$ is the sample mean of original samples $x_1, x_2, \ldots, x_N$;

4) Repeat steps 2) to 3) for B times (e.g., $B = 10{,}000$), where $b = 1, 2, \ldots, B$, obtain the 95% Bootstrap-t CI of t_b^*. If the 95% Bootstrap-t CI does not cover $t_0 = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}}$, then the null hypothesis is rejected at 0.05 level.


FDA
U.S. FOOD & DRUG
ADMINISTRATION