



STATISTICAL ASSESSMENT OF ABUSE-DETERRENT OPIOID DRUG PRODUCTS

LING CHEN, PH.D.

FDA/CDER/OTS/OB/DBVI



Disclaimer

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

Outline

- Abuse potential of drugs and drug scheduling
- Three categories of the assessment of abuse-deterrent (AD) opioids
- Statistical assessment of AD opioids in clinical abuse potential studies
 - Design and statistical analysis of clinical abuse potential studies
 - Comparison between a proposed AD product and an approved AD version of the same opioid product
- Remarks

Why Do People Like to Take Drugs of Abuse?

- The brain is structured to respond to positive changes in our environment.
- When our brain experiences something that is rewarding, there is a tendency to want to go back to repeat that experience (in other words, the experience is considered to produce positive reinforcement).
- Historically, reward was thought to be mediated by the acute release of dopamine in the brain (especially since all drugs of abuse initially increase dopamine).

Continued...

- But decades of neuroscience research now shows that dopamine is neither necessary nor sufficient to produce reward, even when drugs of abuse are administered!
- Instead, the research shows that dopamine release increases the **association the brain makes** between taking the drug, cues associated with the drug and receiving a good outcome from the drug (feeling rewarding sensations).
- This means dopamine can increase simply in anticipation of receiving a drug of abuse because of that association.
- This is the neural basis of a person being motivated to take their drug of choice again.

Abuse Potential - How do we define it?

- *Drug Abuse*

- Defined as the **intentional, non-therapeutic** use of a drug product or substance, even once, to achieve a desired psychological or physiological effect.

- *Abuse potential*

- Refers to the likelihood that abuse will occur with a particular drug product or substance with CNS activity.

- ❖ Desired psychological effects can include euphoria, hallucinations and other perceptual distortions, alterations in cognition, and changes in mood.

The Controlled Substances Act (CSA)

(see generally 21 U.S.C. 811)

- It is the statute establishing federal U.S. drug policy under which the manufacture, importation, possession, use, and distribution of certain substances is regulated.
- The CSA contains five schedules of control: Schedules I, II, III, IV and V.
- Drugs or other substances with a high abuse potential, no currently accepted medical use, and a lack of accepted safety for use under medical supervision are controlled in Schedule I.
- Drugs or other substances with abuse potential that do have a currently accepted medical use (e.g., the drug or substance is in an FDA-approved product) are placed into Schedule II, III, IV, or V.
- The specific placement of a drug or other substance within Schedules II-V is determined by the relative abuse potential of the drug or substance and the relative degree to which it induces psychological or physical dependence (21 U.S.C. 812(b)).

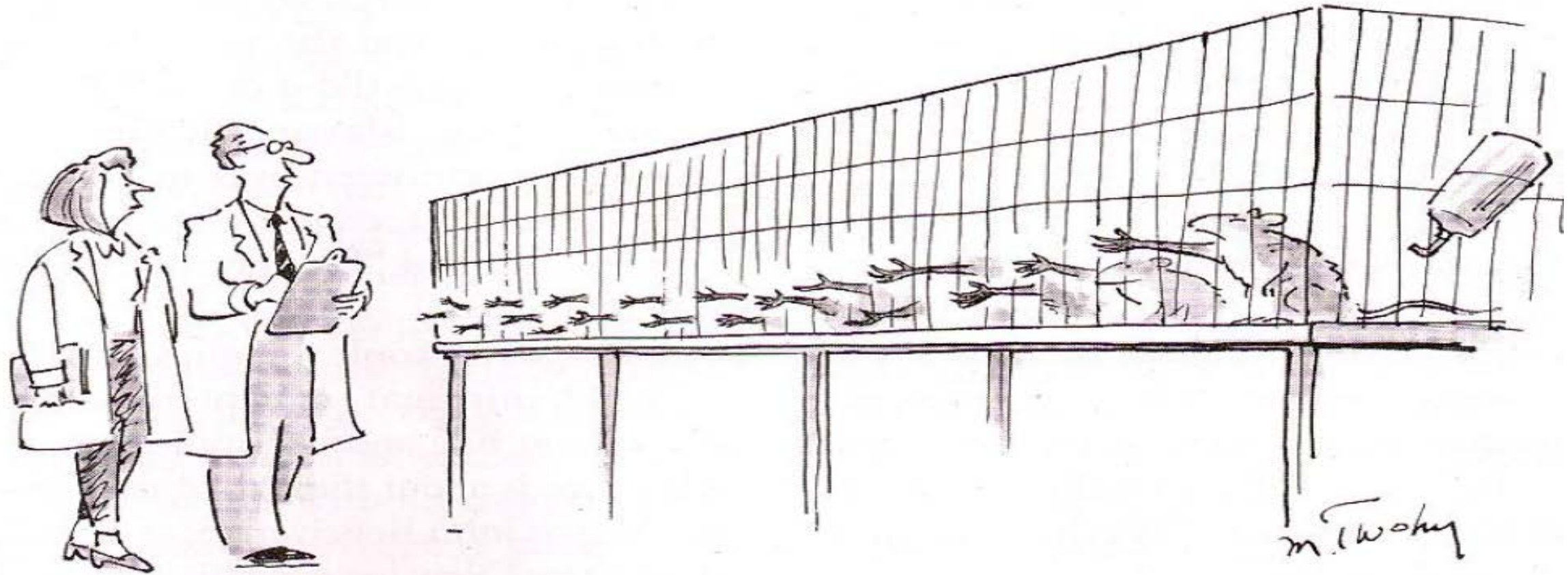
Criteria for Scheduling and Schedules under the CSA

| | | | | | |
|--|---|--|--|---|--|
| C R I T E R I A | Abuse Potential | | | | |
| | High | High | Low relative to CII | Low relative to CIII | Low relative to CIV |
| | No Medical Use | Medical Use | | | |
| S C H E D U L E S | Psychological or Physiological Dependence | | | | |
| | Lack of accepted safety under medical supervision | Severe Psych or Physical | High Psych or Moderate to low Physical | Ltd Psych or Physical relative to CIII | Ltd Psych or Physical relative to CIV |
| | SCHEDULE I | SCHEDULE II | SCHEDULE III | SCHEDULE IV | SCHEDULE V |
| | Heroin Hallucinogens Marijuana Others * A Schedule I substance can be studied under IND | Opioids Barbiturates Cocaine Amphetamine Methylphenidate Methamphetamine PCP | Opioids (Codeine combinations, Buprenorphine) Barbiturates (combinations and products) Ketamine GHB Marinol Anabolic Steroids | Benzodiazepines and other depressants (Zaleplon, Zolpidem, Eszopiclone) Fenfluramine Modafinil Butorphanol Tramadol | Opioids in limited quantities and in combinations (Codeine, Dihydrocodeine, Difenoxin) Pregabalin Lacosamide |

Abuse Potential Assessment

- *Abuse potential assessment*
 - Is science based and requires a comprehensive evaluation of :
 - Chemistry (receptor binding studies)
 - Animal pharmacology studies (in vitro and in vivo)
 - Pharmacokinetics and pharmacodynamics studies
 - Human abuse potential studies
 - Epidemiological studies
- *Based on the totality of data from all studies, the Controlled Substance Staff (CSS) at the FDA will decide whether scheduling recommendation is needed.*

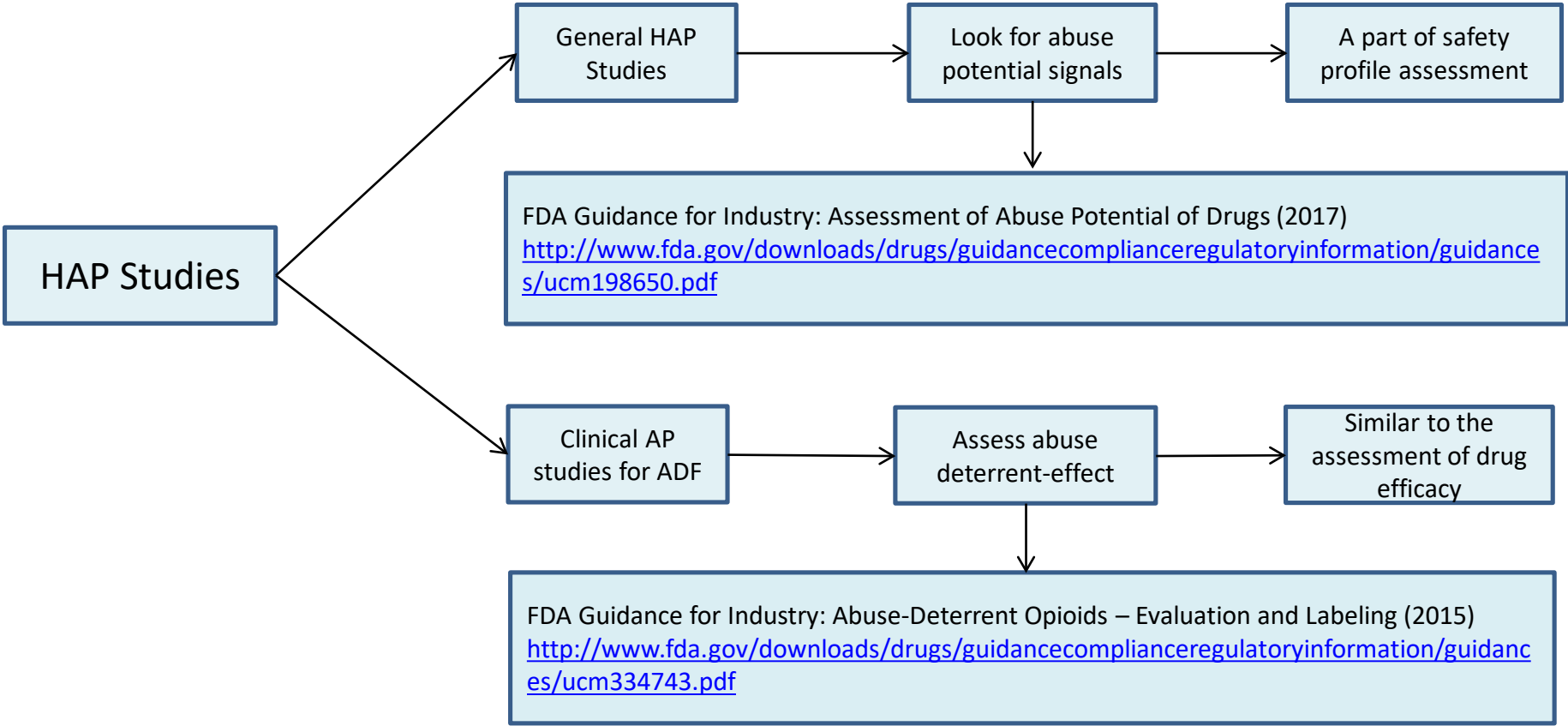
Abuse Potential Assessment



“At this point, we know it’s addictive.”



Two Types of Human Abuse Potential (HAP) Studies



ADF – Abuse-Deterrent Formulation

Opioids

- Every day, more than 115 people in the United States die after overdosing on opioids.
- In 2015, more than 33,000 Americans died as a result of an opioid overdose, including prescription opioids, heroin, and illicitly manufactured fentanyl, a powerful synthetic opioid.
- That same year, an estimated 2 million people in the United States suffered from substance use disorders related to prescription opioid pain relievers, and 591,000 suffered from a heroin use disorder (not mutually exclusive).

Reference

1. CDC/NCHS, National Vital Statistics System, Mortality. CDC Wonder, Atlanta, GA: US Department of Health and Human Services, CDC; 2017.
2. Center for Behavioral Health Statistics and Quality (CBHSQ). 2015 National Survey on Drug Use and Health: Detailed Tables. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2016.

- Opioids fall into schedules across the entire Controlled Substances Act: Schedule I (heroin) – high abuse potential, no approved medical use.
 - Schedule II (oxycodone, hydrocodone)
 - Schedule III (buprenorphine, codeine combinations)
 - Schedule IV (dextropropoxyphene)
 - Schedule V (low dose codeine)



What Does It Mean for an Opioid Drug Product to Be Abuse-Deterrent?

- Opioids with abuse-deterrent properties are designed to deter abusers from using the product for the purpose of abuse.
- Opioid analgesics are commonly manipulated for abuse. Tablets or capsules may be crushed for snorting or dissolved for injection. Extended-release opioid analgesics tablets or capsules may also be crushed to cause defeat of the extended-release characteristics for oral ingestion.
- Abuse-deterrent (AD) products are intended to either make such manipulation more difficult, for example, by making a tablet difficult to crush or for the crushed material to gel upon exposure to water, or include an opioid antagonist to reduce or block the rewarding effects of the opioid (for example, Embeda is an agonist/antagonist combination of morphine and naltrexone).

Three Categories of Premarketing Studies

- Category 1: Laboratory-based in vitro manipulation and extraction studies
- Category 2: Pharmacokinetic (PK) studies
- Category 3: Clinical abuse potential studies
- The results of Category 1 studies may influence the design of PK studies and clinical abuse potential studies by suggesting the methods of manipulation that would yield the greatest release of opioid.
- The results of Category 2 PK studies may influence the need for Category 3 clinical abuse potential studies, and the designs and goals of these studies. For example, if the extended-release characteristics of an abuse-deterrent opioid formulation cannot be defeated and the PK profile remains unchanged following oral or nasal administration of the manipulated product, oral and intranasal studies of abuse potential may not be necessary.

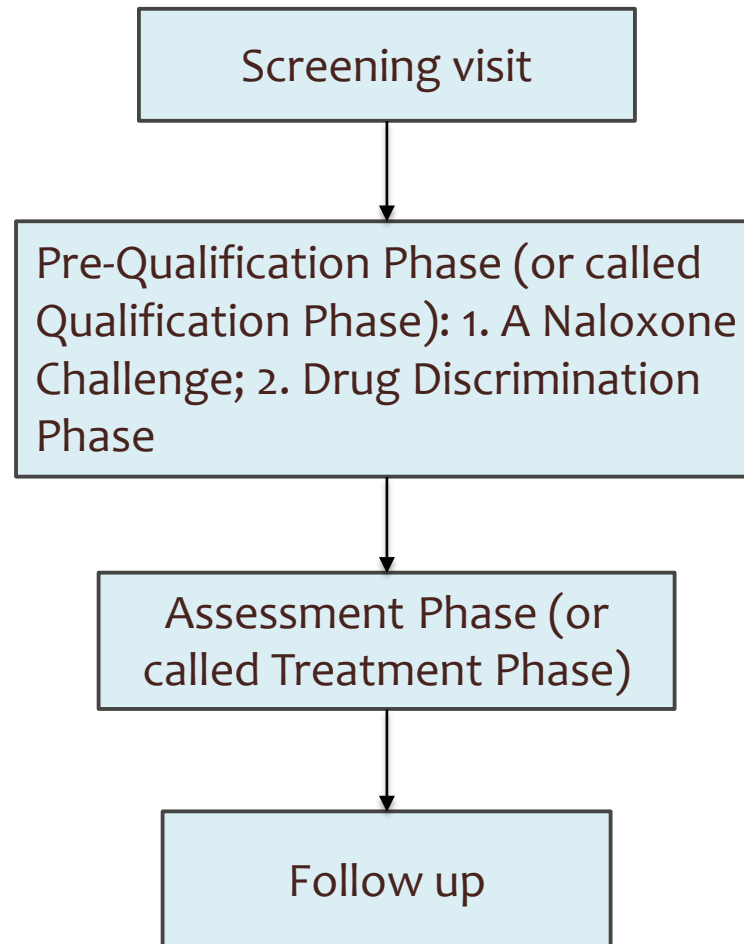
The determination of abuse deterrence of an AD formulation is based on the totality of data from all studies.

DESIGN AND STATISTICAL ANALYSIS OF CLINICAL ABUSE POTENTIAL STUDY

Study Subjects in Clinical Abuse Potential Studies

- The study subjects are not patients. They are opioid-experienced non-dependent subjects who have experience with the particular route of abuse being studied.

Design of Clinical Abuse Potential Studies



A Naloxone Challenge

- A Naloxone Challenge is performed to ensure that the subject is not physically dependent on opioids.
- Only subjects who do not display signs and symptoms of opioid withdrawal as assessed by a Clinical Opiate Withdrawal Scale (COWS) can enter the Drug Discrimination Phase.

Drug Discrimination Phase

- In the Drug Discrimination Phase, subjects will take a conventional immediate release (IR) formulation of the same opioid being developed in an AD formulation and placebo in a blind and crossover manner.
- The route of administration in the Drug Discrimination Phase is the same as that planned for the Treatment Phase.
- The purpose of this phase is to ensure subjects who enter the Treatment Phase can distinguish the conventional IR formulation of the opioid from placebo.
- Therefore, clinical abuse potential study has an enrichment design.

Treatment Phase

- It is usually designed as a randomized, double-blind, placebo- and positive-controlled, crossover study with a repeated Williams Square design.
- For example: 4 treatments crossover, A, B, C, and D denote treatments.

Sequence\Period 1 2 3 4

1 A D B C

2 B A C D

3 C B D A

4 D C A B

Properties of a Williams square design

1. It is a variance balanced design

2. It balances the first-order-carryover effects

- Each subject is randomly assigned to one of treatment sequences. That is, each subject takes all treatments in the study but at different time period. – Each subject is his/her own control.
- There is a washout period between two treatments.

Treatments in the Treatment Phase

- The number of treatments in the Treatment Phase varies.

For example: An typical intranasal study includes (but is not limited to):

1. Manipulated positive control (IR or Non-AD extended release (ER) opioid product)
2. Manipulated test drug (proposed AD product), and
3. placebo.

- If the manipulated test drug has much larger volume than positive control, we suggest that the sponsor include two placebos in the study, one matching the volume as the positive control, and the other matching the volume of the test product.

- The study may also include a lower dose of positive control, which can provide context for the magnitude of the deterrent effect; or include a same dose positive control but matched volume of the test product. This permits evaluation of AD effects not due to volume.

The Primary and Key Secondary Subjective Measures

Primary

Drug Liking Visual Analog Scale (VAS) – 0-100 Bipolar Scale

Question: “Do you like the drug effect you are feeling now?”

0 = “Strong disliking” ; 50 = “Neither like nor dislike”; 100 = “Strong liking”

It is an “at the moment” measure, and data are collected at multiple time points after a single dose administration

Key secondary measure

Take Drug Again (VAS) -0-100 Bipolar Scale

Question: “Given the opportunity, would you take this drug again?”

0 = “Definitely would not”; 50 = “Do not care”; 100 = “Definitely would”

Data are collected at hours 12 and 24 after dosing.

Endpoints

- Pharmacodynamic parameters
 - Emax – Maximum (Peak) Effect
 - TEmax – Time of Peak Effect
 - AUE0-x – Area Under the Effect Curve from 0 to x hours post-dosing, where x = 0.5, 1, or 2 hours
- Primary endpoint
 - Drug Liking Emax
- Key secondary endpoint
 - Take Drug Again Emax

Note: The clinical abuse potential study must show significant reduction for the test product compared to the positive control on the key secondary endpoint to support the approval of the AD labeling.

Quantifying the Abuse-Deterrent Effect

- Define the relative AD effect on maximum liking for a test product compared to its positive control as follows:

$$\theta = \frac{\mu_C - \mu_T}{\mu_C - 50}$$

where μ_C and μ_T denote the means of positive control and test drug, respectively.

Note: The θ is the ratio of the mean difference between the positive control and the test product to the mean of the positive control on the bipolar liking scale, and the mean of the positive control, μ_C , is greater than 50.

The Primary Hypothesis

- The primary hypothesis for Drug Liking Emax is as follow:

$$H_0: \theta \leq \delta^* \text{ vs. } H_a: \theta > \delta^* , \quad (1)$$

where $0 < \delta^* < 1$. If the test result is statistically significant, one may conclude that the test product has a larger than $\delta^* 100\%$ reduction in mean of maximum liking compared to the positive control. The δ^* should be pre-specified in the protocol. The nominal type I error rate is 0.025.

- The hypothesis (1) is equivalent to

$$H_0: \mu_C - \mu_T \leq \delta_1 \text{ vs. } H_a: \mu_C - \mu_T > \delta_1 \quad \text{where } \delta_1 = \delta^*(\mu_C - 50). \quad (2)$$

$$H_0: \mu_T - (1 - \delta^*)\mu_C \geq 50\delta^* \text{ vs. } H_a: \mu_T - (1 - \delta^*)\mu_C < 50\delta^* \quad (3)$$

Hypothesis for the Validation Test

- For the sensitivity and validity of the study, the following hypothesis should be tested at the significance level of 2.5%.

$$H_0: \mu_C - \mu_P \leq 15 \text{ vs. } H_a: \mu_C - \mu_P > 15,$$

where μ_P denotes the mean of placebo.

- One of the selection criteria for a qualified subject in the Qualification Phase is as follows: at least 15-point difference in maximum liking between positive control and placebo during the first 2 hours following drug administration. Therefore, a test value of 15 is required for the validation test.

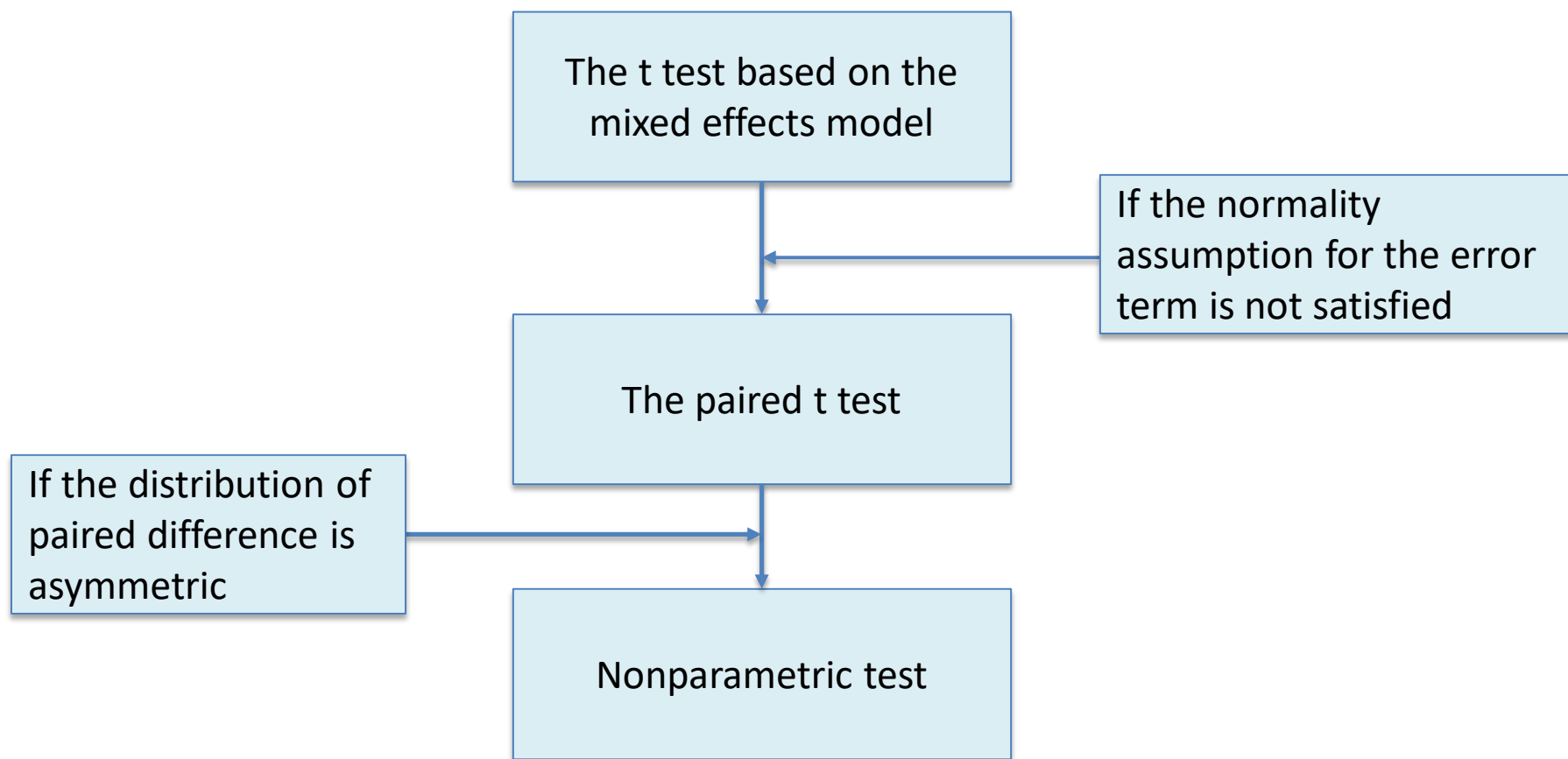
Mixed Effects Model Used in Clinical Abuse Potential Studies

$$y_{ijk} = \mu + \nu_i + \pi_j + \tau_{d[i,j]} + s_{ik} + e_{ijk}$$

Overall mean μ
 Sequence effect ν_i
 Period effect π_j
 Treatment effect $\tau_{d[i,j]}$
 Random subject (sequence) effect s_{ik}
 Random (within-subject) error e_{ijk}

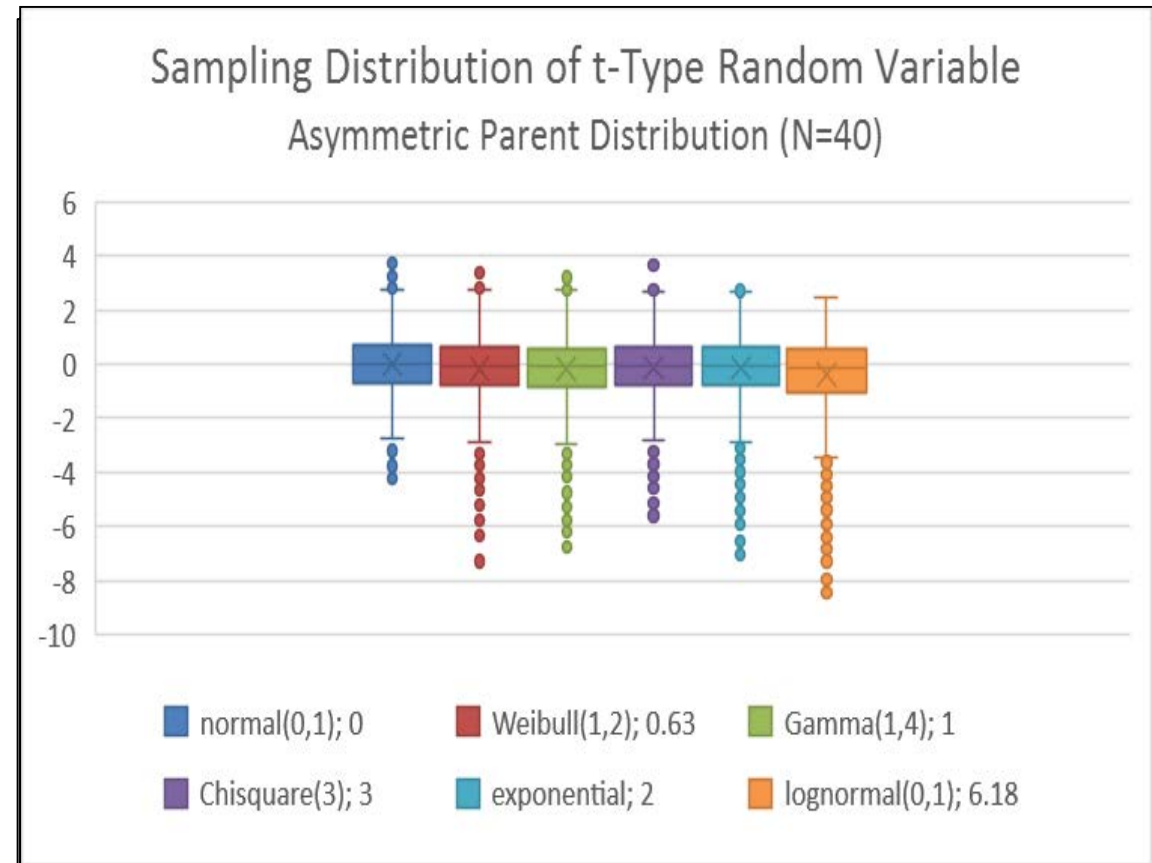
$s_{ik} =_d N(0, \sigma_s^2)$ iid, and $e_{ijk} =_d N(0, \sigma_e^2)$, iid.
 Independent

Statistical Methods



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

- 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. The Sign test is often used for testing the median of difference between two treatments, when t test cannot be used. 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.
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)

About the Unknown δ^*

- The δ^* plays a critical role in the statistical analysis. It is the test value of the hypothesis in (1), should be clinically meaningful and pre-specified in the protocol.
- To avoid finding the sampling distribution of the test statistic for the hypothesis in (1), we use the hypothesis in (3):
- The test can start with the prespecified and clinically meaningful δ^* . If the result from the test for H_0 in (3) is statistically significant, then test the null with a 0.05 increment on δ^* until an insignificant result is obtained. By using this closed testing procedure, one may obtain a larger δ^* than the prespecified one without inflating the type I error rate. One may refine δ^* to a 0.01 increment for the testing procedure.
- Note that the final δ^* is lower than but close to the 97.5% LCL of θ in (1). The closed testing procedure should be prespecified in the protocol, and taken it into consideration in the sample size calculation.

Regarding Sample Size Calculation

- I have seen the following proposal:

Propose a δ^* , and find a historical sample mean of the positive control $\hat{\mu}_C$. Calculate $\hat{\delta}_1 = \delta^*(\hat{\mu}_C - 50)$, and then treat it as a constant to calculate the sample size for the following hypothesis:

$$H_0: \mu_C - \mu_T \leq \hat{\delta}_1 \quad \text{vs.} \quad H_0: \mu_C - \mu_T > \hat{\delta}_1 . \quad (4)$$

- The hypothesis in (4) is not equivalent to the hypothesis in (2), hence is not equivalent of the hypothesis in (1). The proposed method will overestimate the sample size for the study.

Example 1

Table 1. Summary Statistics for Drug Liking Emax.

| Study | N | $\hat{\mu}_C$ | $\hat{\sigma}_C$ | $\hat{\mu}_T$ | $\hat{\sigma}_T$ | r_{CT} |
|-------|----|---------------|------------------|---------------|------------------|----------|
| 1 | 44 | 77.7 | 14.3 | 65.5 | 11.7 | 0.52 |
| 2 | 27 | 94 | 21.1 | 79 | 14.4 | 0.5 |
| 3 | 28 | 90 | 12 | 60 | 12.1 | 0.14 |

Plots of sample size versus δ^* for 3 Studies

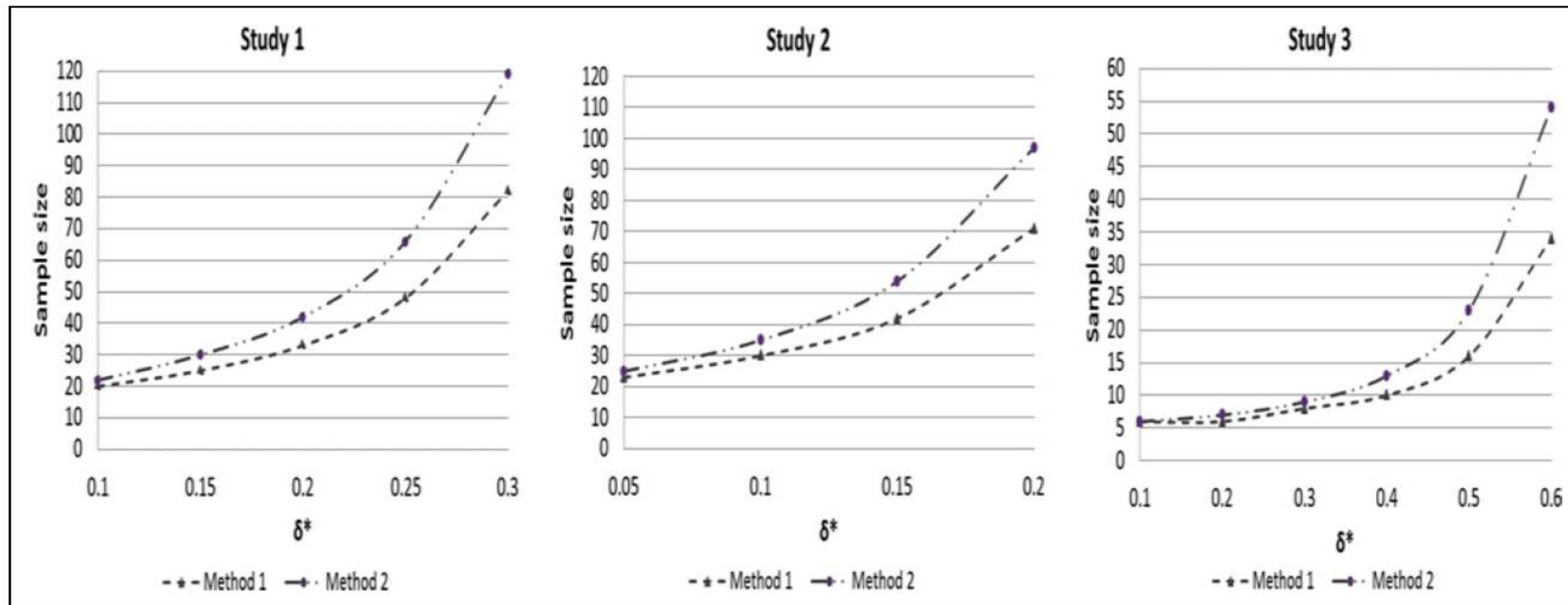


Figure 1 plots sample size versus δ^* for studies 1, 2, and 3. Methods 1 and 2 in the figure denote the sample size calculation based on the tests for the hypotheses in (1) and (4), respectively. In the sample size calculation, the type I error rate $\alpha = 0.025$ and power 0.90.

Secondary Analysis on Drug Liking Emax



- The secondary analysis should be performed on the percent reduction for the potentially abuse-deterrent test product (T) relative to positive control (C) from each individual study subject for Drug Liking VAS on a bipolar scale from 0 to 100.
- One proposed definition of the percent reduction for individual subjects is as follows:

$$\% \text{ reduction} = \frac{C - T}{C - P} \times 100\%,$$

where C, T and P denote Emax scores from an individual subject for positive control, test drug and placebo, respectively.

- It is not a good definition of the percent reduction.

Continued ...

| | C | T | P | % reduction |
|--|----|----|----|--|
|  | 85 | 75 | 75 | $\frac{85-75}{85-75} \times 100\% = 100\%$ |
|  | 85 | 75 | 52 | $\frac{85-75}{85-52} \times 100\% = 30\%$ |

Therefore, a more appropriate definition of percent reduction can be derived by replacing P by the neutral score 50 on a bipolar scale, that is,

$$\% \text{ reduction} = \frac{C - T}{C - 50} \times 100\%,$$

where we assume that $C > 50$. In the case some subjects have $C \leq 50$, define % reduction = 0.

Continued ...

- For penalizing subjects with large placebo responses, the final definition of individual percent reduction is

$$\% \textit{reduction} = \left\{ \begin{array}{l} \frac{C - T}{C - 50} \times \left(1 - \frac{P - 50}{50} \right) \times 100\%, \quad \textit{if } P > 55; \\ \frac{C - T}{C - 50} \times 100\%, \quad \textit{if } P \leq 55. \end{array} \right\}$$

Responder Analysis

- A responder is defined as a subject who had at least $\delta^* 100\%$ of reduction in E_{max} for T relative to C . To ensure that a majority of subjects are responders, a proportion test is recommended to test the null hypothesis that 50% or fewer subjects are responders. That is, test

$$H_0: p^* \leq 50\% \quad vs. \quad H_a: p^* < 50\% \quad (5)$$

at the 2.5% significance level, where p^* denotes the percentage of responders.

- One may use a cutoff point $\delta^* 100\%$ for a responder in the order from small percent to large percent with 5% increment to define a responder, and then test the null hypothesis in (5); that is, the majority subjects are not responders, until an insignificant result is obtained. The $\delta^* 100\%$ also can be refined to a 1% increment in the testing procedure.

COMPARISON BETWEEN A PROPOSED AD PRODUCT AND AN APPROVED AD VERSION OF THE SAME OPIOID PRODUCT

Background

- In April 2015, FDA issued the final guidance to assist pharmaceutical companies in developing opioid products with AD properties. Since 2010, several opioids with AD formation technology have been approved by the FDA, and more are in development.
- The question has been raised regarding how to compare a second generation AD product (a test product) to an approved AD version of the same opioid product (an approved AD product) in a clinical abuse potential study.
- There have been proposals not to include an IR or Non-AD ER opioid product as a positive control in the study, and compare a test product to an approved AD product using a non-inferiority test.

The Reasons for Not Supporting the non-inferiority test

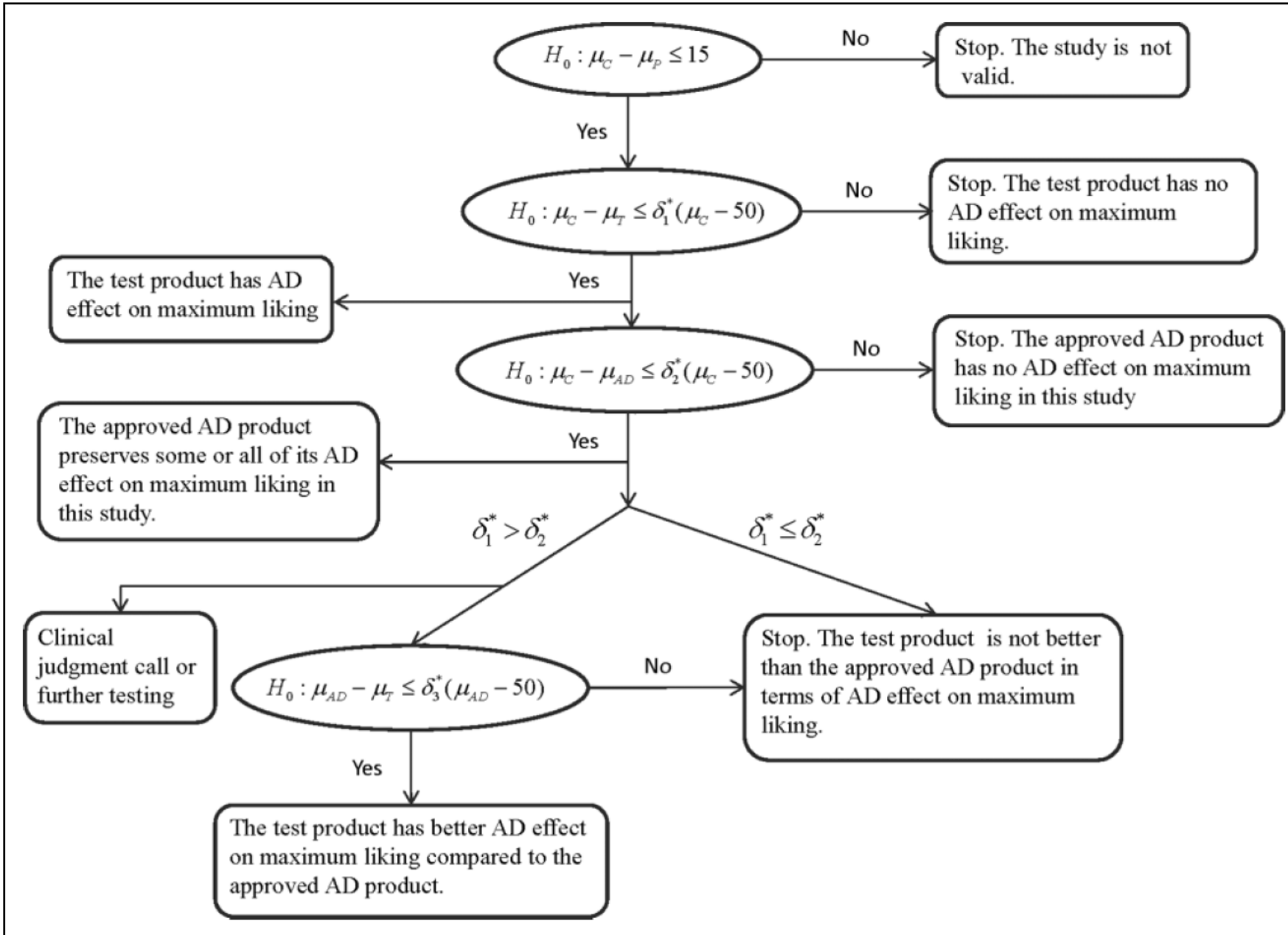
- Suppose the comparison between a test product and an approved AD product is a primary interest of the study. An IR or Non-AD ER opioid product should be included in the study as a positive control, because the study needs
 - to compare the positive control to placebo for the study validation, and
 - to compare the approved AD product to the positive control for examining whether the approved AD product still maintains its AD effect in this study.
- If the validation test fails, the study fails. If the test for the comparison between the approved AD product and the positive control fails, the comparison between the test product and the approved AD product is not meaningful.

Continued ...

- Because IR or Non-AD ER opioid product will be included as a positive control in the study, the second generation of the AD product can be compared to the positive control directly to assess AD property of the product.
- The comparison between the second generation AD product and the first generation of the approved AD version of the same opioid product can be done after the approved AD product demonstrates its AD effect in this study.
- The test for this comparison should be a superiority test, simply because the second generation of AD product should be better than the first generation of AD product. We do not want to put any additional opioid product on the market without better AD properties.



Gatekeeping Testing Procedure



Note: μ_C ; μ_T ; μ_P and μ_{AD} denote the means of positive control, test product, placebo, and approved AD product, respectively. The “Yes” and “No” represent whether the test result is statistically significant or not, respectively. The nominal type I error rate for each test is 0.025.

Example 2

- For demonstrating the gatekeeping testing procedure, data from 2 clinical abuse potential intranasal studies were combined. The sample size and the data were modified to mask these studies. The following Table shows the summary statistics:
- Sample size=28.

| Treatment | Mean | SD | Min | Q1 | Med | Q3 | Max |
|-----------|------|-------|-----|-------|------|-------|-----|
| AD | 76.9 | 20.14 | 36 | 54.75 | 77.5 | 99.75 | 100 |
| C | 90.0 | 10.36 | 61 | 83.5 | 90 | 100 | 100 |
| P | 51.3 | 4.17 | 42 | 50 | 50 | 51 | 67 |
| T | 68.1 | 17.18 | 33 | 51.25 | 66 | 81.5 | 100 |

Abbreviations: AD, approved AD product; C, positive control; P, placebo; SD, standard deviation; T, test product.

Continued...

- The statistical model used in the primary analysis was a linear mixed-effects model that included sequence, period, and treatment as fixed effects and subject(sequence) as a random effect.
- The model assumption for homogeneity of variance was examined using Levene test. The result of the Levene test was statistically significant ($p=0.0001$). Therefore, the heteroskedasticity was adjusted in the model.
- The model assumption of the normality of error term was also examined using Shapiro-Wilk W test on the residuals. The test was not statistically significant with a P -value of 0.0773.

Least Square Means for Drug Liking Emax (N=28)

| Treatment | LSmean | SE | 95% CI | |
|-----------|--------|------|--------|-------|
| | | | LCL | UCL |
| AD | 76.9 | 3.83 | 69.23 | 84.48 |
| C | 90.0 | 2.01 | 86.01 | 93.99 |
| P | 51.3 | 0.81 | 49.64 | 52.86 |
| T | 68.1 | 3.25 | 61.60 | 74.54 |

Abbreviations: AD, approved AD product; C, positive control; CI, confidence interval; LCL, lower confidence limit; P, placebo; SE, standard error; T, test product; UCL, upper confidence limit.

Summary of Results From Statistical Analysis for Drug Liking Emax (N=28)

| Comparison | Test Value/ δ^* | Estimate* | SE | One-Sided P Value | 95% CI for Mean Diff | |
|-------------|------------------------|-----------|------|-------------------|----------------------|------|
| | | | | | LCL | UCL |
| C versus P | 15 | 38.8 | 2.16 | <.0001 | 34.4 | 43.1 |
| C versus T | 0.37 | 11.4 | 3.49 | .0222 | 14.3 | 29.5 |
| C versus AD | 0.11 | 3.2 | 4.23 | .0210 | 4.5 | 21.8 |
| AD versus T | 0 | 8.8 | 5.02 | .0422 | -1.2 | 18.8 |

Abbreviations: AD, approved AD product; C, positive control; CI, confidence interval; P, placebo; SE, standard error; T, test product.

*Estimate of the left-hand side of the inequality in the null hypothesis.

Remarks

- The clinical abuse potential study is an important part of the assessment of AD opioids.
- The δ^* in the primary hypothesis is the test value for the relative AD effect of the test product compared to the positive control, θ . It should be clinically meaningful and prespecified in the protocol.
- One may use a closed testing procedure to obtain a larger δ^* than the prespecified one. However, the closed testing procedure should be pre-specified, and taken into consideration in the sample size determination.
- It is not proper to use a non-inferiority test for the comparison between a test product and an approved AD product. Because an IR or NonAD ER opioid product is in the study, one can compare the test product to an IR or NonAD ER opioid product directly using a superiority test as the primary assessment of the product.

Continued ...

- If the comparison between test product and an approved AD product is also one of the primary objectives, a gatekeeping testing procedure should be used.
- The hypotheses and testing procedures discussed in this presentation are not restricted to Drug Liking VAS. They could be extended to other abuse potential measures.



Acknowledgement

- Silvia N. Calderon, Ph.D. , Senior Pharmacologist, Controlled Substance Staff, Office of Center Directors, CDER, FDA.
- Katherine Bonson, Ph.D., Pharmacologist, Controlled Substance Staff, Office of Center Directors, CDER, FDA.

Reference

1. FDA Guidance for Industry: Abuse-Deterrent Opioids – Evaluation and Labeling, April 2015.
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.
11. Chen, L. (2017). “Principles and Methods of Statistical Assessment of Abuse-Deterrent Opioids,” *Therapeutic Innovation & Regulatory Science*,” Volume 52 Issue 1, 7-13.