Statistical Considerations in OGD's New Draft Guidance for Assessing Adhesion with Transdermal Delivery Systems and Topical Patches for ANDAs

Wanjie Sun, Stella Grosser, Carol Kim, Sam G. Raney

Disclaimer: The opinions and information in this poster are those of the authors, and do not represent the views and/or policies of the U.S. Food and Drug Administration.

FDA/CDER

Abstract

Objective: The statistical approach recommended in previous FDA/CDER product-specific guidances for generic transdermal delivery systems products (TDS) had a low power for "well-adhering" TDS, impacting availability of generic TDS to the public. In 2016, CDER published a new draft guidance to resolve this issue.

Methods: Mathematical proof and simulation were used to evaluate the cause of the low power of the previous statistical non-inferiority (NI) approach. A new NI hypothesis was proposed and a NI margin was determined based on simulation and real data analyses. Power was compared among different approaches.

Results: Regarding the cause for the low power of the historical statistical approach for well-adhering TDS, one common consideration was that the non-normality of the adhesion data (i.e., the high skewness and the discrete feature of ordinal scores) violated the normality assumption of the linear mixed model used to evaluate NI, and therefore, might cause the low passing rate for well-adhering TDS products. Extensive statistical research revealed that non-normality of the adhesion data was not actually the true cause of the problem. Rather, it was determined that the direction of the adhesion scale (a smaller score indicates better adhesion: 0 for perfect adhesion, 4 for complete detachment) coupled with the use of a ratio of the mean (ROM) scores (Test/RLD) for the NI statistical test were the main causes for the low power (passing rate) of the historical statistical approach for well-adhering TDS products.

A new statistical hypothesis was recommended in the draft guidance on adhesion by replacing the traditional ratio-of-means NI test with a difference-of-means (DOM) NI test, still based upon mean adhesion scores. DOM NI test is robust in power to the direction of adhesion scores (whereas ROM NI test is highly sensitive to it), and can dramatically improve the power for "well-adhering" TDS products. The NI margin of 0.15 for the difference of means was determined to be appropriate based upon collaborative research by clinicians, statisticians and other scientists.

Conclusion: The currently recommended statistical approach in the new guidance corrects the low power of the historical statistical approach recommended in the previous product-specific guidance and significantly reduces the needed sample size for well-adhering TDS products while retaining the targeted type 1 error rate under 0.05. It is also consistent with previously passing TDS.

Introduction and Motivation

Adhesion study (cross over or matched parallel) needs to demonstrate that the adhesion performance of the TEST TDS is non-inferior to the RLD TDS.

Historical Statistical Method in Product-Specific Guidance:

$$H_0: \frac{\mu_T}{\mu_R} > 1.25;\ H_1: \frac{\mu_T}{\mu_R} \le 1.25$$

μ_T and μ_R are group means of mean adhesion score averaged across individual scores at each assessment time point for each subject (primary endpoint). NI is established if the one-sided 95% upper confidence bound (UB) of $\mu_T - 1.25\mu_R \le 0$.

Adhesion Scale: 5-point (Recommended by the FDA Guidance): the smaller, the better

0 (≥90% attached), 1 (75-<90%), 2(50-<75%), 3 (>0-<50%), 4 (complete detachment).

Problem: Advances in adhesive technology and TDS design have increased the prevalence of well-adhering products. The historical statistical method has a *Low Power* for *well-adhering* TDS drug products, requiring a very large sample size for adhesion studies. When the RLD mean is close to 0, the required sample size can be thousands.

Table 1. Ratio of Means NI Test: a Hypothetical Well-Adhering TEST TDS Product with Lower (Better) Mean Adhesion Score than the RLD that Fail NI

N	TEST Mean ± SD	RLD Mean ± SD	95% UB of $\mu_T - 1.25\mu_R$	NI Test
80	0. 034± 0.10	0.045 ± 0.12	0.01 (>0)	Fail NI

Objectives

1. To investigate what caused the low power of the historical statistical non-inferiority approach.
2. To develop an alternative non-inferiority approach which can correct the low power of the historical statistical approach.
3. To provide a non-inferiority margin for the new statistical approach.

Methods

Without loss of generality, under normality,

- For a *Ratio of Means* NI test $H_0: \frac{\mu_T}{\mu_R} > \theta;\ H_1: \frac{\mu_T}{\mu_R} \le \theta$,

$$Power = \Phi\left(Z_\alpha - \frac{\mu_T/\mu_R - \theta}{\sqrt{((1-\theta)^2 CV_s^2 + (1+\theta^2) CV_e^2)/2n}}\right)$$

where CV_e and CV_s are the intra-subject and inter-subject CV, n is the number of subjects.

- With a fixed mean ratio μ_T/μ_R, under the alternative hypothesis $H_1: \mu_T/\mu_R = k < \theta$, power is a *monotone increasing* function of the REF mean μ_R: $\frac{\partial Power}{\partial \mu_R} > 0$.
- For a *Difference of Means* NI test $H_0: \mu_T - \mu_R > \theta; H_1: \mu_T - \mu_R \le \theta$, with a fixed mean difference μ_T-μ_R under $H_1: \mu_T$-$\mu_R = k < \theta$, power is *independent* of the REF mean μ_R:

$$Power = \Phi\left(Z_\alpha - \frac{\mu_T - \mu_R - \theta}{\sigma_D/\sqrt{n}}\right), \text{ and } \frac{\partial Power}{\partial \mu_R} = 0$$

Results

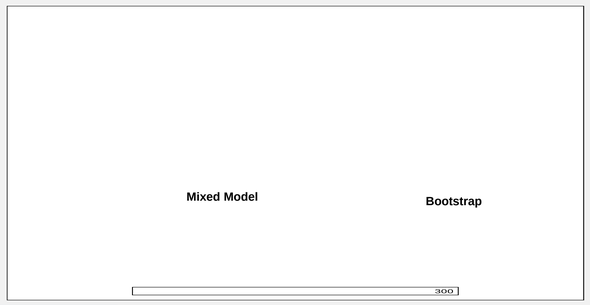
Objective 1: What *caused* the low power of the historical stat approach?

Tentative Answer: It was generally thought that *non-normality (high skewness, ordinal scale)* of the adhesion data causes the low power of the historical statistical approach.

- However, simulation (Figure 1) shows that linear mixed model (parametric) and bootstrap (non-parametric) provide similar power for TDS with a range of adhesion quality (poor to good) when TEST and RLD have equal adhesion means: low power for well-adhering TDS (mean=0.11) and high power for poorly-adhering TDS (mean=0.7).

Conclusion: Non-normality is *NOT* the true cause of the lower power of the historical statistical approach. Otherwise, a non-parametric method (e.g., a bootstrap approach) would have corrected the low power of the historical statistical approach.

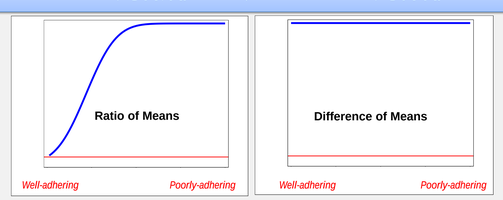
Figure 1. Power of Linear Mixed Model (Left) vs. Bootstrap (Right) For Well-Adhering (Upper) and Poorly-Adhering (Lower) TDS When $\mu_T = \mu_R$



Final Answer:

- As shown in the Methods, for a ratio of means NI test, with a fixed mean ratio (e.g. $\mu_T/\mu_R = k$), power is a monotone *increasing* function of the RLD mean (Figure 2 Left).
- Therefore, for a ratio of means NI test, TEST products with a *larger* REF mean (poorly-adhering based on the current scale) tend to have a *higher* power; whereas TEST products with a *smaller* REF mean (well-adhering) tend to have a *lower* power.
- This explains the low power of historical statistical approach for well-adhering TDS.

Figure 2. Power vs. REF Mean μ_R with a Fixed Effect Size under H_1 For a ROM NI Test ($H_1: \mu_T/\mu_R = k < \theta$) and a DOM NI Test ($H_1: \mu_T$-$\mu_R = k' < \theta'$)



Therefore, *the direction of the adhesion scale* (where a smaller score indicates better adhesion: i.e., 0 for perfect adhesion, 4 for complete detachment) coupled with the use of a *ratio of the mean scores* for the NI test, rather than *non-normality*, were the true cause for a low passing rate of the historical statistical approach when the RLD was a well-adhering TDS (and a high passing rate when RLD TDS was poorly adhering).

Objective 2: Develop An Alternative Statistical Approach

Aspects Evaluated:

- Alternative Measure of Treatment Effect
- Alternative Adhesion Scales
- Alternative Statistical Methods (Parametric and Non-parametric)

Selected Statistical Approach:

A new statistical hypothesis was recommended by replacing the traditional *ratio of means* NI test with a *difference of means* NI test using mean or weighted mean adhesion scores averaged across assessment time points throughout patch wear based on the current 5-point adhesion scale because of the following considerations:

- The new approach greatly *improves power (Figure 3) and reduces needed sample size (Figure 4)* for well-adhering TDS while retaining the targeted type 1 error rate under 0.05 (Sun *et. al.* 2017).
- A difference of means NI test is *robust in power* regardless of the direction of scores (Figure 2 Right) or other commonly seen location shifts, whereas a ratio of means NI test is not (Figure 2 Left) – power remains consistent for difference of means but changes vastly for ratio of means whether using the current scale or reversed scale.
- A difference of means NI test is *robust* to *adjustment* for study design *covariates* whereas a ratio of means NI test depends upon the adjusted value of covariates.
- The new approach has *minimum impact* on the current long-standing adhesion scale.

Objective 3: What NI Margin Should be Used?

The NI margin of 0.15 for the difference of means was determined to be appropriate based upon collaborative research by clinicians, statisticians and other scientists.

Final Statistical Approach Adopted in the New Guidance

- In the new draft guidance, a new statistical hypothesis was recommended by replacing the traditional *ratio of means* NI test with a *difference of means* NI test using mean adhesion scores across time based on the current 5-point scale, and a NI margin of 0.15 was recommended for the difference of means.

Old: $\boldsymbol{H_0: \frac{\mu_T}{\mu_R} > 1.25;\ H_1: \frac{\mu_T}{\mu_R} \le 1.25}$

New: $\boldsymbol{H_0: \mu_T - \mu_R > 0.15;\ H_1: \mu_T - \mu_R \le 0.15}$

Figure 3. Power of Ratio of Means (Left) vs. Difference of Means (Right) for Well-adhering (Upper) and Poorly-adhering (Lower) TDS When $\mu_T = \mu_R$ Using Current Scale

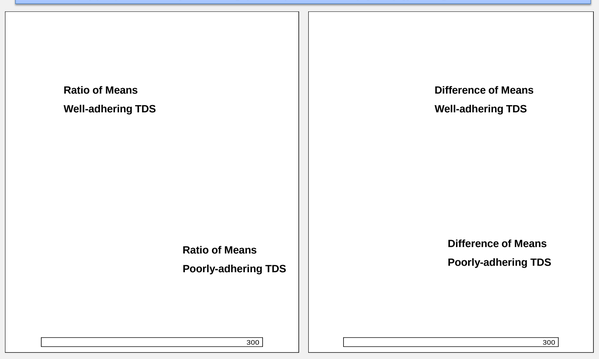
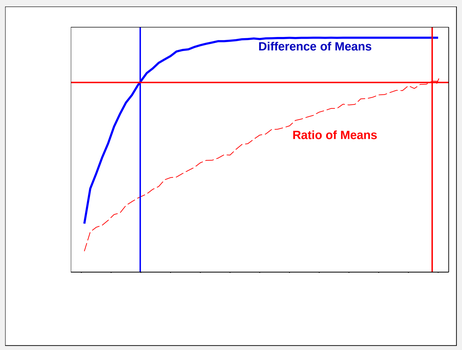


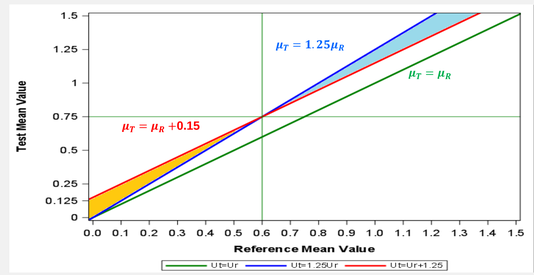
Figure 4. Reduction of Sample Size from Ratio of Means NI Test to Difference of Means NI Test for a Hypothetical Well-adhering TDS with TEST and RLD Mean = 0.26, Intra-subject SD=0.6



Revisit Table 1. A Hypothetical Well-Adhering TEST TDS Product with Better Adhesion than the RLD

NI Hypothesis	TEST Mean ± SD	RLD Mean ± SD	95% UB	NI Test
Old: $\boldsymbol{H_1: \frac{\mu_T}{\mu_R} \le 1.25}$	0. 034± 0.10	0.045 ± 0.12	$UB\ of\ \mu_T - 1.25\mu_R =$ 0.01 (>0)	*Fail NI*
New: $\boldsymbol{H_1: \mu_T - \mu_R \le 0.15}$	0. 034± 0.10	0.045 ± 0.12	$UB\ of\ \mu_T - \mu_R =$ 0.015 (<0.15)	*Pass NI*

Figure 5. A Schematic Diagram for Comparison between Old (Blue) vs. New (Red) Hypothesis Assuming No Data Variation



- In Figure 5, the area below the *blue* line establishes NI according to the *old* guidance: $\frac{\mu_T}{\mu_R} \le 1.25$, and the area below the *red* line establishes NI according to the *new* guidance: $\mu_T - \mu_R \le 0.15$.
- The *green* line is when TEST and RLD are *identical*: $\mu_T = \mu_R$, which is *parallel* to the *red* line - the new *non-inferiority* criteria . The *same* criteria (0.15 of absolute mean difference) is applied for *all* products, whether well-adhering or poorly-adhering.
- The *yellow* shade represents the region (RLD mean < 0.6) where power is gained under the new guidance / recommendation.
- The *blue* shade represents the loss of excess power for poorly-adhering products (RLD mean >0.6) under the new guidance / recommendation.

Conclusions

- *Direction* of adhesion scale (a smaller score for better adhesion) coupled with the use of a *ratio of the mean* NI test, rather than *non-normality*, were the main cause for the low passing rate of the historical statistical approach when the RLD TDS adhered well.
- The *direction* of the scoring scale can impact power inappropriately when using a NI test based upon a *ratio of means. In particular,* With a fixed mean ratio $\boldsymbol{\mu_T/\mu_R}$, under the alternative hypothesis $\boldsymbol{H_1: \mu_T/\mu_R = k > \theta}$, power is a *monotone increasing* function of the REF mean $\boldsymbol{\mu_R}$ – the lower $\boldsymbol{\mu_R}$ (better adherence), the lower the power.
- However, the *direction* of scores does *NOT* impact the power with a *difference of means* NI test, which is also robust to other location shifts of scales (e.g., reversing the scale, or scale + a constant).
- The recommended new statistical approach *corrects the low power* of the NI test for well-adhering TDS while retaining the targeted type 1 error rate under 0.05.
- The new statistical approach *significantly reduces the needed sample size for well-adhering TDS products*, which reduces the burden on applicants.
- The new guidance is also consistent with previously passing TDS.

Impact on Generic Drug Review

- Since publication of the new guidance in June 2016, several well-adhering generic TDS products have *passed NI* for adhesion by using the new statistical approach recommended in the new guidance, which otherwise would *fail* to pass NI if using the historical statistical method.

References

- FDA Draft Guidance - Assessing Adhesion with Transdermal Delivery Systems and Topical Patches for ANDAs, June 2016
- Sun W, Grosser S, Tsong Y. Ratio of means vs. difference of means as measures of superiority, non-inferiority, and average bioequivalence, Journal of Biopharmaceutical Statistics, 2017, 1-18.