# A Comparison of the European Medicines Agency (EMA) and United States Food and Drug Administration (US FDA) Approaches to Assessing/Comparing In Vivo Adhesion for Transdermal Delivery Systems (TDS)


**Wanjie Sun[1], Stella Grosser[1], Carol Kim[2], Sam G. Raney[3]**

[1]US FDA/Center for Drug Evaluation and Research (CDER)/Office of Translational Sciences/Office of Biostatistics/Division of Biostatistics VIII, [2]US FDA/CDER/Office of Generic Drugs (OGD)/Office of Bioequivalence/Division of Clinical Review, [3]US FDA/CDER/OGD/Office of Research and Standards/Division of Therapeutic Performance



## Abstract

The EMA and US FDA approaches to evaluating TDS adhesion are fundamentally similar. Both are based upon using a clinical study to estimate the percentage of the TDS surface area that remains adhered at progressive points in time, with an expectation of non-inferior adhesion performance for a Test (T) compared to Reference (R) product. The recommended EMA and US FDA study designs are also similar. Both typically use the highest strength TDS for the labeled duration of wear, discourage reinforcement of the TDS, and accommodate a design that may also be used to evaluate pharmacokinetic endpoints. The few key differences relate to technical matters, like the use of a one- vs. two-component statistical approach, a primary endpoint using the last assessment vs. all assessments, the scales used to estimate TDS adhesion, how the temporal profile of TDS adhesion influences the assessment (and whether scores are carried forward), or the acceptable margin for a difference in adhesion performance between a T and R product. This work compares the EMA and US FDA approaches, and discusses areas for harmonization.

The EMA's recommendations relate to new as well as generic TDS products, with an expectation that the 90% confidence interval (CI) of mean adhesion for the T product at the end of the dosing interval should lie above 90%. This criterion corresponds to a one-arm non-inferiority (NI) test: $H_0$: $\mu_T \leq 90\%$; $H_1$: $\mu_T > 90\%$. When a R TDS exists as a reference for comparing the adhesion performance of the T TDS, and < 90% of the area of the R TDS remains adhered at the end of the dosing interval, the lower limit of the 90% CI for the difference in adhesiveness (T-R), using the percentage of adhesion as a continuous variable, should not be less than -10%. This criterion corresponds to a two-arm NI test: $H_0$: $\mu_T - \mu_R < -10\%$; $H_1$: $\mu_T - \mu_R \geq -10\%$.

The US FDA's recommendations focus on generics, and the expectation is that a generic TDS should provide a consistent, uniform adhesion of its entire surface area to the skin for the entire duration of wear. The US FDA's primary endpoint for the adhesion assessment is the (weighted) mean adhesion score, calculated from individual adhesion scores at each assessment, averaged across all time points (not just the last time point at the end of the dosing interval). The individual adhesion scores are based on a 5-point numerical scale, where each score represents a defined range for the percentage of TDS area that remains adhered. The recommended statistical approach evaluates the difference of the (weighted) mean adhesion scores between T and R with an acceptable margin of 0.15. This corresponds to a two-arm NI test: $H_0$: $\mu_T - \mu_R > 0.15$; $H_1$: $\mu_T - \mu_R \leq 0.15$.

Statistical tests, simulations, and meta-analysis are used to compare results based upon these technical differences, and to identify opportunities for alignment that may help to facilitate inter-agency harmonization, establish more consistent global standards, and potentially reduce barriers for the global development of high quality TDS products.


## Introduction

- In November 2014, the EMA published a guideline entitled "Guideline on the pharmacokinetic and clinical evaluation of modified release dosage forms" which provides recommendations on the evaluation of clinical adhesion for TDS products (in Appendix II: in vivo skin adhesion).
- In June of 2016, the US FDA Center for Drug Evaluation and Research published a draft guidance entitled "Assessing Adhesion with Transdermal Delivery Systems and Topical Patches for ANDAs" which also provides recommendations on the evaluation of clinical adhesion for TDS products.

## Objectives

- Compare the EMA and FDA approaches recommended in these two publications, focusing exclusively on the implications for generic TDS products
- Discuss areas for harmonization of EMA and FDA standards on TDS adhesion

## Methods

### Comparison of EMA vs. FDA Study Design Recommendations

- The EMA and US FDA *approaches to evaluating TDS adhesion are fundamentally similar:* both are based upon using a clinical study to estimate the percentage of the TDS surface area that remains adhered at progressive points in time, with an expectation of non-inferior adhesion performance for a Test (T) compared to Reference (R) product.
- The recommended EMA and US FDA *study designs are also similar*. Both typically use the highest strength TDS for the labeled duration of wear, discourage reinforcement of the TDS, and accommodate a design that may also be used to evaluate pharmacokinetic endpoints.
- The few *key differences* relate to *technical* matters, like the *one-component* vs. *two-component* statistical approach, a *primary endpoint* using the last assessment vs. overall mean of all assessments, the *scales* used to estimate TDS adhesion, how the *temporal profile* of TDS adhesion influences the assessment (and whether scores are carried forward), or the acceptable *margin* for a difference in adhesion performance between a T and R product.

### The Current EMA Recommended Statistical Approach

- Primary Endpoint: the adhesion score measured as the *percentage* of area that remains *adhered at the end of the dosing interval* for each TDS product.
- EMA's recommended statistical approach has *two components*:

*Criterion 1 (C1)*: "In general, it is expected that the 90% CI of mean adherence for T at the end of the dosing interval [$\mu_T$] should lie above 90%." Under FDA's framework of hypothesis testing, operationally, C1 corresponds to *a one-group one-sided test at* $\alpha$ *=5%*:

$$H_0: \mu_T \leq 90\%;\ H_1: \mu_T > 90\%$$ (*higher* value is better)

*Criterion 2 (C2)*: "If it is considered unlikely that this requirement can be met it may be possible to establish NI of T to R. The may be possible if R has poor adherence (<90%). The lower limit of the 90% CI for the [mean] difference of adhesiveness [at the end of the dosing interval] (test [$\mu_T$] - reference [$\mu_R$], using the percentage of adhesion as a continuous variable, should not be < -10%." Similar to C1, operationally, C2 corresponds to *a two-group NI test at* $\alpha = 5\%$:

$$H_0: \mu_T - \mu_R < -10\%;\ H_1: \mu_T - \mu_R \geq -10\%$$ (*higher* value is better)

### The Current US FDA Recommended Statistical Approach

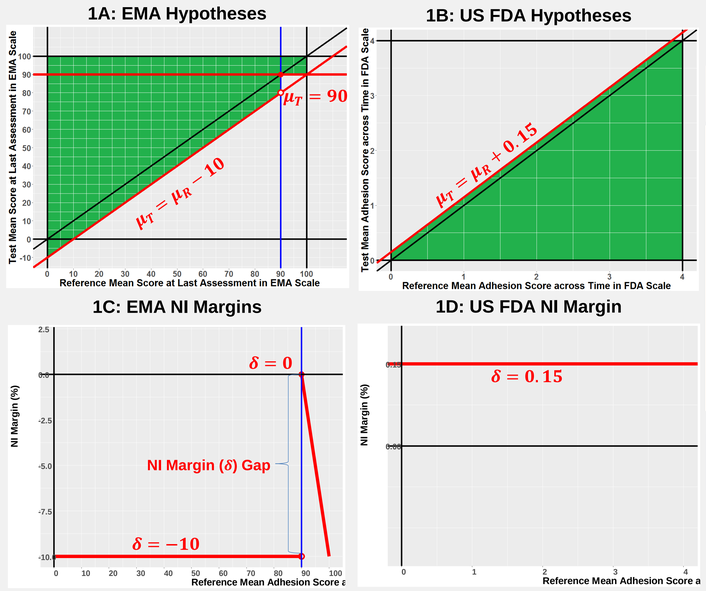
- Primary Endpoint: the *(weighted) mean adhesion score* for each TDS, calculated from individual adhesion scores at each assessment averaged *across all time points*, after applying *worst observation carried forward* (WOCF), based on the widely-used *5-point* adhesion scale:

0 (90 to100% adhesion), 1 (75 to <90%), 2 (50 to <75%), 3 (>0 to <50%), 4 (0%)

- US FDA's recommended statistical approach is *a two-group NI hypothesis test* based upon the difference of (weighted) mean adhesion scores between T ($\mu'_T$) and R ($\mu'_R$) with an acceptable margin of 0.15:

$$H_0: \mu'_T - \mu'_R > 0.15;\ H_1: \mu'_T - \mu'_R \leq 0.15$$ (*lower* value is better)

Figure 1. Schematic Comparison of EMA vs. US FDA NI Hypotheses (1A & 1B) and Corresponding NI Margins (1C & 1D): Green is Establishment of NI (Rejecting $H_0$)




- For generics, *EMA*'s two-component statistical approach may introduce a *discontinuity* in the NI margin when the R product exhibits an adhesion of 90% at the end of the dosing interval.
  - When the adhesion of the R product is *<90%* (e.g. 89.9%), Criterion 2 is an *NI* test (NI margin =-10%)
  - When the adhesion of the R product is exactly *90%*, Criterion 1 may be interpreted to be a *superiority* test (NI margin = 0)
  - When the adhesion of the R product is *>90%*, Criterion 1 is an *NI* test where the NI margin progressively decreases from 0 to -10% as the R mean adhesion increases from 90% to 100%.
  - This apparent discontinuity is shown in the *triangle* (grey) in Figure 1A when the R mean is between 90% and 100%: power is *lowest* when the R mean is exactly *90%* at the last assessment.
- For generic drugs, *US FDA*'s statistical approach is *consistently an NI test* with a *constant NI margin* (0.15) for all TDS products, regardless of the adhesion performance of the R product.

### Simulations and a Meta-Analysis are Used to Compare the Power and NI Results for Generic TDS Adhesion Studies Using the EMA and US FDA Approaches

*Simulations:* generated from multivariate normal distributions (MVN) for 100% scale adhesion data. Intra-subject correlation across different time points with the same treatment group and between T and R for the same subject are incorporated by the variance-covariance matrix of MVN

## Simulation Results

Figure 2. EMA vs. US FDA: Similar Power for Well- Adhering (Top Panels) and for Poorly-Adhering (Bottom Panels) TDS

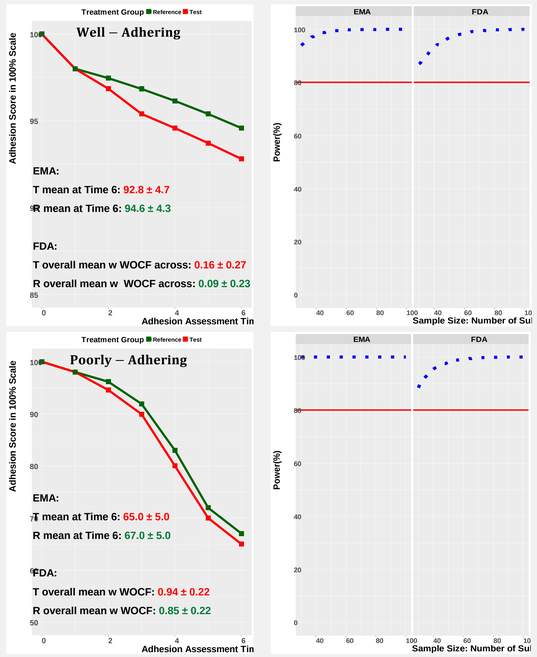



Figure 3. EMA vs. US FDA: Discontinuity in Power for EMA Approach when R Mean at Last Visit is 90% (T and R have Identical Distribution)

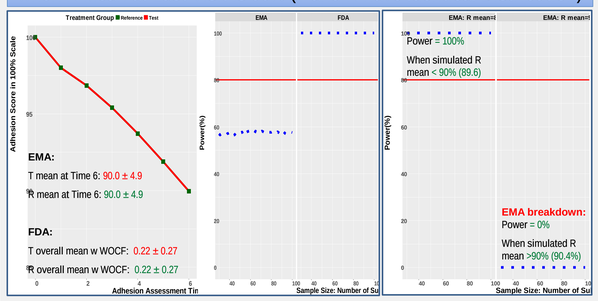
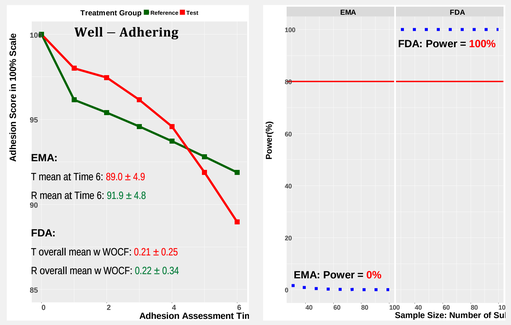



Figure 4. EMA vs. US FDA: Impact of Different Temporal Profiles on the Power of the EMA vs. US FDA Approaches




- T and R are both well-adhering TDS products (generally > 90% adhered), with similar overall mean adhesion scores across all time points, but each has different mean adhesion scores at the last assessment.
- The EMA approach does not establish NI, but the US FDA establishes NI

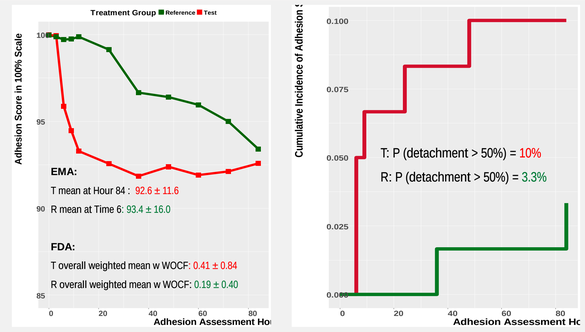
***Disclaimer:*** **The opinions and information in this poster reflect the views of the authors and should not be construed to represent U.S. Food and Drug Administration's views or policies.**

## Meta-Analysis of Results

Out of 9 studies with TDS adhesion observations measured using a 100% scale:

- 6 studies had consistent outcomes (for well adhering and moderately-well-adhering TDS) using either the EMA or US FDA approaches: The 6 studies demonstrated that the adhesion performance of the T TDS was non-inferior to that of the R TDS in each case.
- 3 studies had discrepant outcomes: In these 3 studies, the adhesion performance of the T TDS was only found to be non-inferior to that of the R TDS based upon the EMA approach, but not based upon the US FDA approach.

Figure 5. Hypothetical Study: Key Features Mimic Studies with Discrepant NI Outcomes for EMA and US FDA Approaches in the Meta-Analysis




| | EMA | | | | US FDA | | |
|---|---|---|---|---|---|---|---|
| N | T vs R Mean ± SD Score at Last Assessment | R Mean Score at Last Assessment | EMA Criterion 1 (C1) or 2 (C2) | Pass NI | T vs R Overall Weighted Mean ± SD Score with WOCF | UB of $\mu_T - \mu_R$ | Pass NI |
| T: 60 R: 60 | T: 92.6 ± 11.6 R: 93.4 ± 16.0 | R: 93.4% > 90% | EMA C1: Lower Bound of CI for T = 90.1% (>90%) | Yes | T: 0.41±0.84 R: 0.19±0.40 | 0.42 >0.15 | No |

## Conclusions

*An analysis of the results suggests that*

- The *EMA and US FDA approaches* to assessing/comparing TDS adhesion *are fundamentally similar* in terms of the general approach, the clinical study design considerations, and ~~similar~~ acceptance criteria based upon a demonstration of statistical non-inferiority.
- Simulations and a meta-analysis demonstrated *consistent outcomes by EMA and US FDA approaches* in most of the studies evaluated, involving well-adhering, and moderately well-adhering TDS.

*Potential Areas for Harmonization Between the EMA and the US FDA*:

- The EMA approach has the benefit of addressing TDS adhesion for both new & generic products. A harmonized inter-agency standard would benefit from, similarly, defining standards for both new & generic products, including potential reformulations of the new product (for which a R TDS would exist).
- A consequence of the EMA's *two-component* approach for new and generic products is that it appears to create a *discontinuity* in the NI margin (Figure 1) and power (Figure 3) for generic drugs, effectively imposing a *superiority standard* when the R TDS adhesion is 90% at the last assessment time, and an *NI test with a margin close to 0* when the R TDS adhesion is slightly >90%.
- The US FDA's approach has the benefit that it is *consistently an NI test* with a *constant NI margin*, regardless of the adhesion performance of the R TDS. A similarly consistent approach for generic (or reformulated) TDS may be worthy of consideration for a harmonized inter-agency standard.
- The EMA's primary endpoint has the benefit of simplicity, using the *last assessment* only. However, the US FDA's primary endpoint has the advantage that it can discriminate TDS products which have comparable adhesion at the last assessment, but *different temporal profiles* for adhesion during the product wear (Figure 5). The relative benefits of these different approaches to defining the primary endpoint deserve further consideration.
- The EMA's approach to measuring TDS adherence has the benefit that it can estimate the area adhered more *precisely* than the US FDA's 5-point scale. However, FDA's 5-point scale may be more *sensitive* to differentiating differences in detachment of a magnitude that may be clinically significant. The relative benefits of these different approaches to estimate the area of the TDS that remains adhered deserve further consideration.