HFUX RESEARCH
medical human factors

Why $H_0: ER_T - ER_R > d$ $H_A: ER_T - ER_R \leq d$ ***?***

When you can $x^2 = \Sigma \frac{(O-E)^2}{E}$ ***!***

It's Hip to be Square!

Demonstrating Equivalency without Inferiority in CUHF Studies

Heidi M. Mehrzad
HFUX Research, LLC

Over*VIEW*

- **Human Factors and Usability**
- **The Problem with ANDA**
- **The Elephant in the Room: Noninferiority**
- **Equivalence in Drug Efficacy ≠ Safe and Effective Use**
- **It's so Much More Hip to be Square than to be Inferior**
- **The Chi Square Trifecta: ANDA, the Real World, and Innovation**

Human Factors *and* Usability

FDA Website (2024)

“Why is Human Factors Engineering important to medical devices?

For medical devices, the most important goal of the human factors/usability engineering process is to minimize use-related hazards and risks and then confirm that these efforts were successful, and users can use the device safely and effectively.”

HumanFactors AND *USABILITY*

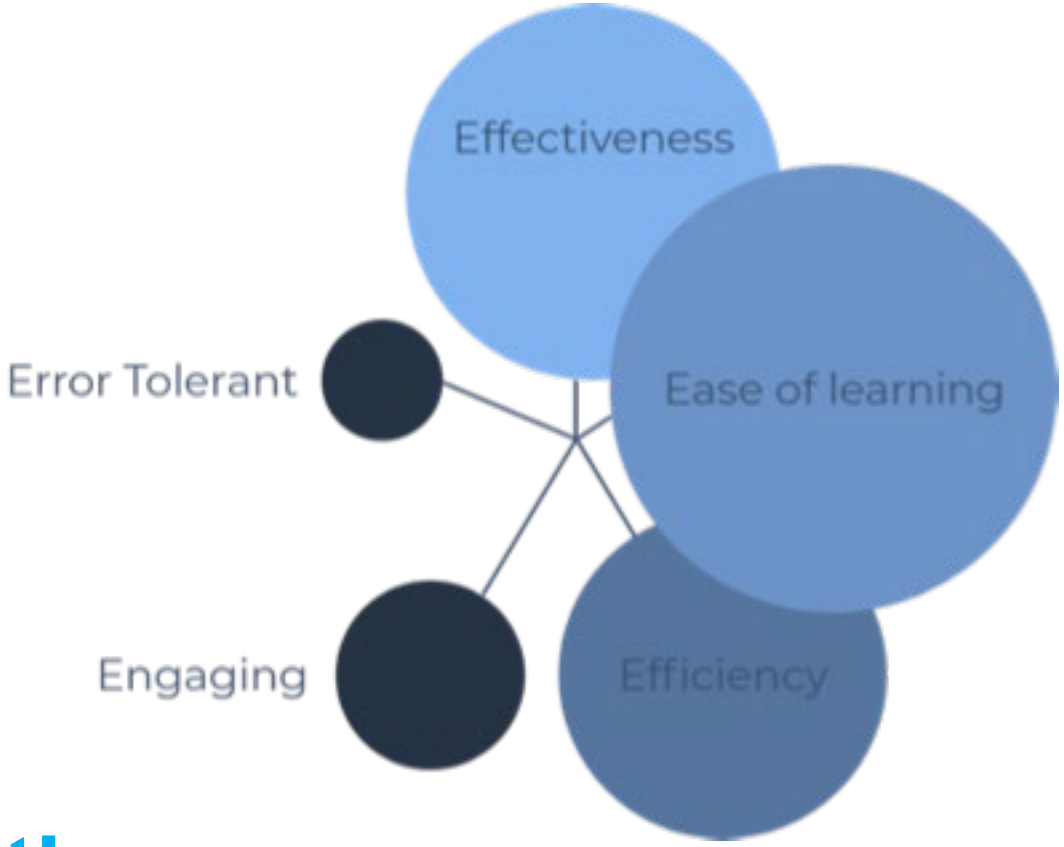
Human Factors is the **science** we apply to **research, design** and **engineer** a product with respect to the **end user**.

Usability is the **goal** and **measure** used when **evaluating** these efforts with respect to the **end user**.

Measuring USABILITY THE 5 Es

The 5Es

- Effective
- Efficient
- Engaging
- Error Tolerant
- Easy to Learn



Goal: Safe and Effective

In Summary…

What is human factors?

Human Factors is the science we apply to research, design and engineer a product with respect to the end user.

What is human factors in drug-device combination product development?

Human Factors is a Risk-Based Approach.

What are its goals?

To minimize use-related hazards and risks related to use of the product and mitigate such through the design of the device UI.

How do we measure it?

Through usability evaluation of the UI of the device with focus on its: effectiveness, efficiency in use, user engagement, its error tolerance, as well as its learnability.

Why do we do it?

To ensure users can safely and effectively use a user interface (design).

FDA ANDA *and* Current Approach

”The Drug Price Competition and Patent Term Restoration Act of 1984 (Pub. L. 98-417) (the Hatch-Waxman Amendments) created, among other things, section 505(j) of the Federal Food, Drug, and Cosmetic Act (FD&C Act). Under section 505(j), **an ANDA applicant can rely on FDA’s previous finding that the RLD is safe and effective so long as the ANDA applicant demonstrates that the proposed drug product and the RLD are the same with respect to active ingredient(s), dosage form, route of administration, strength, and, with certain exceptions, labeling.”**

FDA CDER Guidance (2017) – Comparative Analyses and Related Comparative Use Human Factors Studies for a Drug-Device Combination Product Submitted in an ANDA: Draft Guidance for Industry

Current ANDA GUIDANCE

- Hatch-Waxman Section 505(j) of the FD&C Act:
  - An ANDA applicant can rely on FDA's previous finding that the RLD is safe and effective so long as the ANDA applicant demonstrates that the proposed drug product and the RLD are the same with respect to active ingredient(s), dosage form, route of administration, strength, and, with certain exceptions, labeling.
- **Drug products** that are **approved in ANDAs** are generally considered by FDA to be **therapeutically equivalent** to their RLD.
- **Products classified** as **therapeutically equivalent** can be **substituted** with the full expectation that the **generic product will produce the same clinical effect and safety profile as the RLD** under the conditions specified in the labeling.

”A **generic combination product** classified as **therapeutically equivalent** to the RLD can be **expected to produce the same clinical effect and safety profile as the RLD** under the conditions specified in labeling. **This does not mean, however, that the proposed generic combination product and its RLD need to be identical in all respects. FDA recognizes that an identical design may not always be feasible** and, in certain instances, differences in the design of the user interface for a generic combination product as compared to the RLD may exist without precluding approval of the generic combination product under an ANDA.”

FDA CDER Guidance (2017) – Comparative Analyses and Related Comparative Use Human Factors Studies for a Drug-Device Combination Product Submitted in an ANDA: Draft Guidance for Industry

Identical Design **NOT** *REQUIRED*

- Therapeutically equivalent **does not mean** that the **proposed generic** combination product **and its RLD need to be identical in all aspects**.
- An **ANDA is not required to be the same as RLD and can differ** from it in certain aspects.
- **FDA recognizes** that an **identical design may not always be feasible** and differences in the design of the user interface for a generic compared to the RLD may exist **without precluding approval under an ANDA**.
  - **And while FDA recognizes the possibility of design differences,**
  - **Industry understands they are almost inevitable.**

Industry CUHF CHALLENGES

- **Lack of public data.**
  - Absence of available market data to support estimate of true error rates for RLD.
- **Setting NI margin** (arbitrary?).
- Requirements of **large sample sizes**.
  - Recruitment.
  - Surrogates.
  - Rare and orphaned diseases.
- **Lack of resources**, e.g., expertise in drug efficacy statistics.
- **Rigidity in study design.**
- **Losing sight of true HF risk-based approach.**

Drug Efficacy
≠ Safe *and*
Effective Use

”**FDA does not consider** the **comparative use human factors studies** described in this guidance **to be clinical investigations** intended to demonstrate the safety or effectiveness of the proposed generic combination product. **Rather**, the comparative use human factors studies described in this guidance are **intended to confirm that the differences in device and labeling** between the generic combination product and RLD **are acceptable**, and that the proposed **generic** combination product **can be substituted with** the full expectation that the generic combination product will produce the **same clinical effect and safety profile** as the RLD under the conditions specified in the labeling.”

FDA CDER Guidance (2017) – Comparative Analyses and Related Comparative Use Human Factors Studies for a Drug-Device Combination Product Submitted in an ANDA: Draft Guidance for Industry

Non Inferiority Model

$$H_0: ER_T - ER_R > d$$
$$H_A: ER_T - ER_R \leq d$$

Where:

$\boldsymbol{d}$ $= NI\ Margin\ between\ ER_R\ and\ ER_T$
$\boldsymbol{H_0}$ $= Null\ Hypothesis$
$\boldsymbol{H_A}$ $= Alternative\ Hypothesis$
$\boldsymbol{ER_T} = Error\ Rate\ Generic$
$\boldsymbol{ER_R} = Error\ Rate\ RLD$

Current **NiModel** *APPROACH*

- **Determine the allowable margin (*d*)** by which ER_T could exceed ER_R.
- **Calculate the study sample size** considering assumed error rates and *d*.
- **Observe error rates for the critical task(s)** during the experiment.
- **Perform the statistical hypothesis test**:
  - **H_0: $ER_T - ER_R > d$**
  - **H_A: $ER_T - ER_R \leq d$**
- **Reject or accept null hypothesis.**

” NI tests comparing use error rates with the delivery device constituent part of a proposed generic combination product to those of the RLD **are similar to usual statistical tests for a difference but translated to account for allowable differences in error rates** between the proposed generic combination product and its RLD. **In comparing pharmaceutical products, NI tests are often conducted to indirectly demonstrate that a proposed product is more efficacious than a placebo**. In contrast, **a comparative use human factors study** with an NI design as described in this guidance **is intended to help confirm one aspect of the substitutability of a proposed generic combination product for its RLD, and not for determining differences relative to a placebo.”**

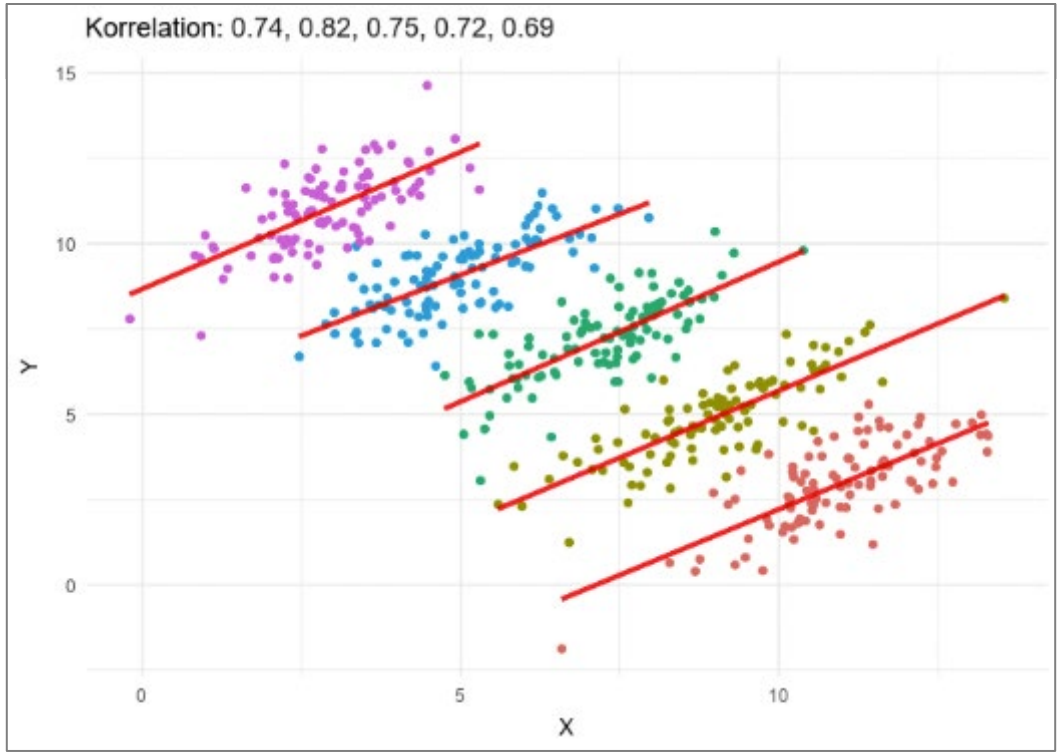
FDA CDER Guidance (2017) – Comparative Analyses and Related Comparative Use Human Factors Studies for a Drug-Device Combination Product Submitted in an ANDA: Draft Guidance for Industry

Drug Efficacy **Does Not Equal** *SAFE and EFFECTIVE use*

- **Distinguish drug efficacy from safe and effective use.**
- Is a drug efficacious - NI Model is a common statistic in clinical investigations/pharmacology.
- What's the behavior associated to safe and effective use?
  - Usability performance - NI Model applicable to this?
  - Is safe and effective use just an artifact in this?
- **NI model focuses only on pass/fail mindset of tasks, i.e., use errors.**
- **NI model does not look at the behavior associated with "safe and Effective use," instead it limits its focus on use errors relating to design differences only.**
- Study design and statistical model should be identified and selected based on the research question you are trying to answer.
- **We should focus more on the study design demonstrating equivalence in safe and effective use and then select a model most befitting.**

Simpson's **Paradox** *VIZUALIZED*

- Simpson's paradox on data resembling real-world variability indicates that risk of misjudgment of true causal relationship can be hard to spot.



POSITIVE TREND

Why not just conduct a Human Factors Validation Study?

”Potential applicants should note that the **objective of a comparative use human factors study differs from the objective of human factors validation studies**. Specifically, **human factors validation studies are not designed to assess differences in use error rates for specific external critical design attributes between two products**. Therefore, the human factors validation report and studies, as described in FDA’s guidance entitled, “Applying Human Factors and Usability Engineering to Medical Devices,” are separate and distinct from the comparative use human factors study...”

FDA CDER Guidance (2017) – Comparative Analyses and Related Comparative Use Human Factors Studies for a Drug-Device Combination Product Submitted in an ANDA: Draft Guidance for Industry

”FDA would generally accept a proposed generic combination product that had the same rates of error as the RLD, as demonstrated by an adequately designed comparative use human factors study or studies.”

FDA CDER Guidance (2017) – Comparative Analyses and Related Comparative Use Human Factors Studies for a Drug-Device Combination Product Submitted in an ANDA: Draft Guidance for Industry

Chi Square Test

$$x^2 = \Sigma \frac{(Observed - Expected)^2}{Expected}$$

Where:

$\mathbf{x^2} = Chi\ Square\ Obtained$

$\mathbf{\Sigma} = Sum\ Of$

$\mathbf{Observed} = Observed\ Value(s)$

$\mathbf{Expected} = Expected\ Value(s)$

Chi SQUARE TEST

- A chi-squared test is a statistical hypothesis test used in the analysis of contingency tables.
- The test is primarily used to examine whether two categorical variables (two dimensions of the contingency table) are independent in influencing the test statistic (values within the table).
- The purpose of the test is to evaluate how likely the observed frequencies would be assuming the null hypothesis is true. The test is valid when the test statistic is chi-squared distributed under the null hypothesis:
  - **Chi-squared test is used to determine whether there is a statistically significant difference between the expected frequencies and the observed frequencies in one or more categories of a contingency table.**

”The sample size of a comparative use human factors study should be adequate to support a demonstration that design differences of a generic combination product do not impact the product’s clinical effect or safety profile compared to the RLD. The sample size required to support a showing that the difference […] is negligible depends on conditions under which the experiment is run.“

FDA CDER Guidance (2017) – Comparative Analyses and Related Comparative Use Human Factors Studies for a Drug-Device Combination Product Submitted in an ANDA: Draft Guidance for Industry

Chi**SQUARE** *STUDY DESIGN*

- **4 User Group Categories**:
  - RLD Experienced: Expose to RLD
  - RLD Experienced: Expose to Generic
  - RLD and Generic Naïve: Expose to RLD
  - RLD and Generic Naïve: Expose to Generic
- **User groups selected based previously identified drug-device intended end users** and corresponding distinct user profiles/groups.
- **Sample size** per distinct user group = *n* = **15, e.g., Faulkner model.**
- **All participants perform 3 doses**.
- **Use Chi-square** test to see if there is a **relationship between two categorical variables**, e.g., use errors and number of attempts for critical tasks, etc.
  - If no significance is found, product can be said to show no differences in use problems; and, if significance is found, product can be said to show differences in use problems.

TASK A n = 15 per group	**RLD**	**Generic**
RLD Using RLD	**X**	
RLD Using Generic		**X**
Naïve to Both Using RLD	**X**	
Naïve to Both Using Generic		**X**

Chi Square allows for more flexibility in study design and greater diversity in subsequent data analyses and correlating statistics.

Chi**SQUARE** *VARIABLES*

- **Independent Variables**
  - Product, RLD vs. Generic
- **Dependent Variables**
  - Critical Tasks
    - Use Error (Counts and Ratio)
    - Time on Task
    - Number of Attempts of Tasks (within one use scenario) (Close Calls?)
    - Number of Attempts of Seeking Clarifying Information (IFU usage?)
- **Co-Variates**
  - User Demographics and Backgrounds
    - Years of Experience
    - Age
    - Dexterity Scale

Chi SQUARE *STATISTICS*

TASK A n = 15 per group	**Observed** Use Errors	**Expected** Use Errors	**TOTAL**
RLD Using RLD	5	6.25	**1.56**
RLD Using Generic	8	6.25	**0.49**
Naïve to Both Using RLD	5	6.25	**1.56**
Naïve to Both Using Generic	7	6.25	**0.09**
	25 25/4 = 6.25		**3.7** (chi square value)

$$x^2 = \Sigma \frac{(O-E)^2}{E}$$

$$RLD_{RLD} = x^2 = \Sigma \frac{(5-6.25)^2}{6.25} = 1.56$$

$$RLD_{Generic} = x^2 = \Sigma \frac{(8-6.25)^2}{6.25} = 0.49$$

$$Naïve_{RLD} = x^2 = \Sigma \frac{(5-6.25)^2}{6.25} = 1.56$$

$$Naïve_{Generic} = x^2 = \Sigma \frac{(7-6.25)^2}{6.25} = 0.09$$

$$\textbf{\textit{Chi Square Value}} = 3.7$$

$$\textbf{\textit{df}} = 3$$

$$\textbf{\textit{p}} = 0.05$$

$$\textbf{\textit{Right Tail Probability}}\ \alpha\ (\text{Table}) = 7.815$$

$$7.815 > 3.7 = \textit{Not Significant} = \textit{No Differences in Use Errors between Categories}$$

Chi SQUARE STATISTICS

TASK A n = 15 per group	Observed Use Errors	Expected Use Errors	TOTAL
RLD Using RLD	5	9.5	**2.13**
RLD Using Generic	14	9.5	**2.13**
Naïve to Both Using RLD	14	9.5	**2.13**
Naïve to Both Using Generic	5	9.5	**2.13**
	38 38/4 = 9.5		**8.52** (chi square value)

$$x^2 = \Sigma \frac{(O-E)^2}{E}$$

$$RLD_{RLD} = x^2 = \Sigma \frac{(5-9.5)^2}{9.5} = 2.13$$

$$RLD_{Generic} = x^2 = \Sigma \frac{(14-9.5)^2}{9.5} = 2.13$$

$$Naïve_{RLD} = x^2 = \Sigma \frac{(14-9.5)^2}{9.5} = 2.13$$

$$Naïve_{Generic} = x^2 = \Sigma \frac{(5-9.5)^2}{9.5} = 2.13$$

$$\textbf{Chi Square Value} = 8.52$$
$$df = 3$$
$$p = 0.05$$
$$\textbf{Right Tail Probability } \alpha \text{ (Table)} = 7.815$$

$$7.815 < 8.52 = \textbf{Significant} = \textbf{Differences in Use Errors between Categories}$$

Chi SQUARE *CELLVALUE < 5*

TASK A n = 15 per group	**Observed** Use Errors	**Expected** Use Errors	**TOTAL**
RLD Using RLD	5	5.5	0.045 = **0.05**
RLD Using Generic	8	5.5	1.136 = **1.14**
Naïve to Both Using RLD	3	5.5	1.136 = **1.14**
Naïve to Both Using Generic	6	5.5	0.045 = **0.05**
	22 22/4 = 5.5		2.38 (chi square value)

$$x^2 = \Sigma \frac{(O-E)^2}{E}$$

$$RLD_{RLD} = x^2 = \Sigma \frac{(5-5.5)^2}{5.5} = 0.05$$

$$RLD_{Generic} = x^2 = \Sigma \frac{(3-5.5)^2}{5.5} = 1.14$$

$$Naïve_{RLD} = x^2 = \Sigma \frac{(3-5.5)^2}{5.5} = 1.14$$

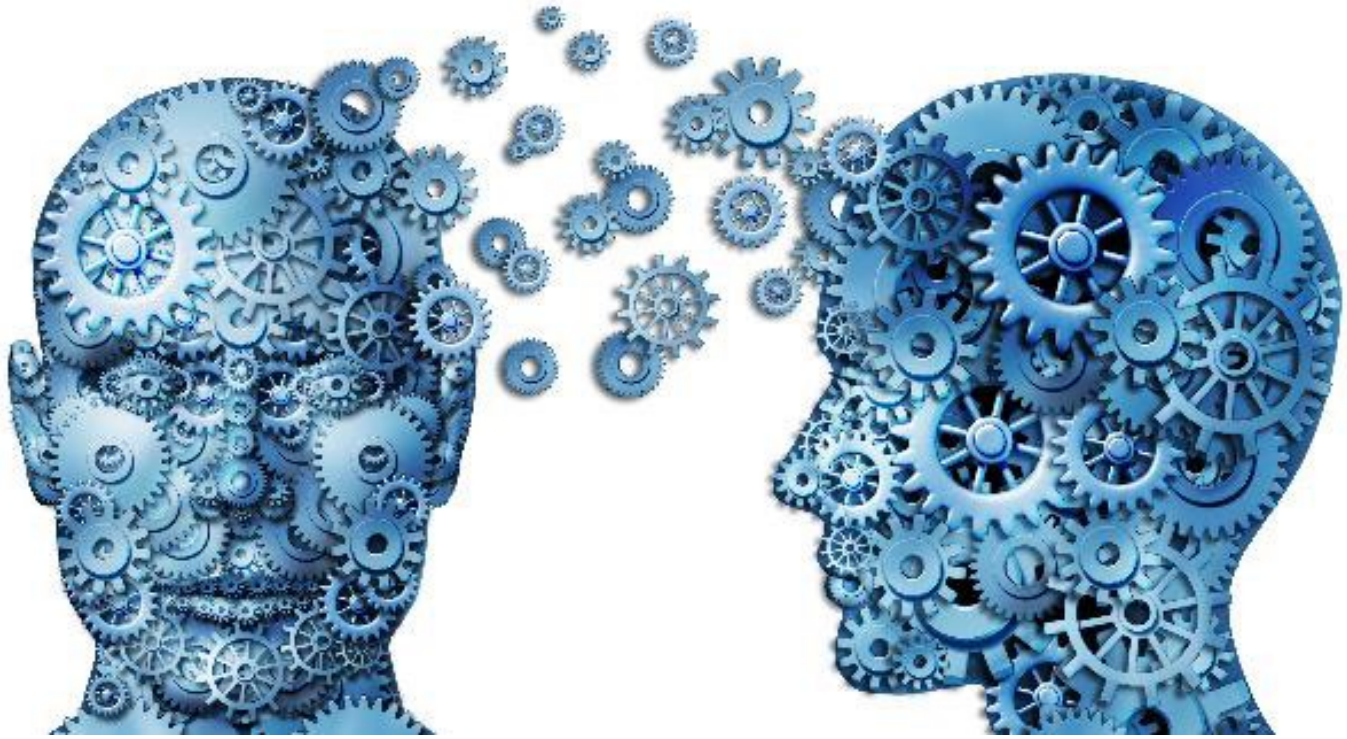
$$Naïve_{Generic} = x^2 = \Sigma \frac{(6-5.5)^2}{5.5} = 0.05$$

What Now?

Fisher's Exact Test

For contingency tables with smaller sample sizes (and cell values less than 5), a Fisher's exact test is used instead.

ANDA *the* Real World *and* Innovation



”FDA recognizes that a potential applicant of a proposed generic combination product may develop a user interface that has certain differences from the user interface approved for the RLD. FDA may accept such design differences if they are adequately analyzed, scientifically justified, and do not preclude approval in an ANDA.”

FDA CDER Guidance (2017) – Comparative Analyses and Related Comparative Use Human Factors Studies for a Drug-Device Combination Product Submitted in an ANDA: Draft Guidance for Industry

Everybody: *ANDA filings do not allow for innovation.*

Me: *Why on earth not?*

ANDA Is ~~Not~~ For *INNOVATION*

- If we **move away from current CUHF study design** and corresponding statistical model that is **rooted in the principles of drug efficacy evaluations**, and **instead focus on comparing actual USABILITY PERFORMANCE with respect to SAFE and EFFECTIVE USE**, we could:
  - **Allow for outdated RLD UIs to be re-imagined in Generics, and**
    - Mitigate RLD existing known-use errors,
    - Adapt Generic UI designs to today's technology standards,
    - Address RLD existing design issues not accounting for neurodivergent and disabled end users, and
    - Address RLD existing design issues not accounting for real-life use scenarios in today's healthcare system and its health insurance and prescription issuing/filling challenges.

Patients: *Depending on my insurance I might use the Generic prior to the RLD.*

Noninferiority: *Say what now?*

Chi Square: *No problem. Let's Go!*

Group **CATEGORY** *INTERCHANGEABILITY*

- The Chi Square Model allows interchangeability of use error differences by tasks between group categories.
  - **RLD vs. Generic**
    - Patients/Users using RLD first and switching to Generic.
  - **Generic vs. RLD**
    - Patients/Users using Generic first and switching to RLD.

Meaning:

- **This model allows statements of “safe and effective use” to account for patients/users not only switching from RLD to Generic, but also from Generic to RLD.**

”In determining the margin d, the **variability in ER_R**, which is an expected observation when conducting an experiment on any product, **should be considered as well as the risk any difference in outcomes will pose to patients**. That is, the value of d will differ between products, **depending on the indication(s) and the clinical consequences associated with failing to perform the critical tasks appropriately.”**

FDA CDER Guidance (2017) – Comparative Analyses and Related Comparative Use Human Factors Studies for a Drug-Device Combination Product Submitted in an ANDA: Draft Guidance for Industry

Root CAUSE *ANALYSIS*

- **Root Cause Analysis**
  - Use Errors
  - Close Calls
  - Use Difficulties
    - Case-By-Case
    - Design Related
    - Residual Risk Analysis
    - Risk Benefits Statement
- **Move away from a model that only considers a pass/fail mind set, i.e., only identifying, evaluating, and analyzing use problems leading to failures (use errors), towards a model that focuses on all use problems encountered during all (critical) tasks to ensure a comprehensive comparison of usability performance between RLD and Generic to enable a robust statement of substitutability from a HF safe and effective criteria.**

Just Kidding…
To Be Continued…

Questions **ANSWERS** *EXAMPLES*

Any questions or device examples you would like to discuss?

Further questions? Email me, I DO love questions – in a normal type of way!

Heidi M. Mehrzad

founder and consultant

HF*UX* Research

info@hfuxresearch.com

www.hfuxresearch.com

Fiscal Year 2024
Generic Drug Science & Research Initiatives
Public Workshop

For more information, please visit:

Coffee Break

We will begin promptly at 3:35 P.M Eastern Time (GMT -4)