BAYESIAN TIER 1 STATISTICS
IABS 2017

JOSÉ G. RAMÍREZ
AMGEN, INC.

AMGEN®
Pioneering science delivers vital medicines™

OUTLINE

- Tier 1 Statistics
- Tier 1 Example
- Effect Size
- Bayesian Tier 1 Statistics
- Summary

FDA (NONBINDING RECOMMENDED) TIERS

250 *2. Determination of the Statistical Methods to be Used*
251
252 FDA's current approach to evaluating analytical similarity is to define three tiers corresponding to the
253 use of three different methods for comparing attributes. FDA believes that the use of these three tiers
254 with appropriate similarity acceptance criteria should help support a demonstration that the proposed
255 biosimilar is highly similar to the reference product. Equivalence testing (Tier 1) is typically
256 recommended for quality attributes with the highest risk ranking and should generally include assay(s)
257 that evaluate clinically relevant mechanism(s) of action of the product for each indication for which
258 approval is sought. The use of quality ranges (Tier 2) is recommended for quality attributes with a
259 lower risk ranking, and an approach that uses visual comparisons (Tier 3) is recommended for quality
260 attributes with the lowest risk ranking. The three methods are described in Section IV.B.

Statistical Approaches to Evaluate Analytical Similarity – Guidance for Industry September 21, 2017

AMGEN

TIER 1: EQUIVALENCE TEST

- **Difference between the sample means $Mean_T - Mean_R$**
- ***Equivalence* is demonstrated within a practical relevant range, or *equivalence acceptance criteria* (EAC).**
- **EAC = $f \times \sigma_R$**
  - **σ_R = standard deviation of the reference product**
  - ***f = 1.5***
    - **"Use of this multiplier in computing the equivalence margin results in a test with reasonable properties under what we feel are realistic conditions"**

Tsong, Y., Dong, C., Shen, M., (2017), "Development of Statistical Approaches for Analytical Biosimilarity Evaluation".

AMGEN

TIER 1: EQUIVALENCE TEST

- **Test conducted using two one-sided t-tests (TOST) or**
  - **To claim equivalence both tests have to be rejected**

- **90% confidence interval around $Mean_T$ – $Mean_R$**
  - **Equivalent if 90% CI entirely within the ±EAC= *1.5* × s_R**

$$-EAC \leq (Lower\ Confidence\ Limit; Upper\ Confidence\ Limit) \leq +EAC$$

TIER 1: EQUIVALENCE TEST

90% confidence interval calculated using

- **Unequal variances**
- **Satterthwaite approximation to degrees-of-freedom or**
- **Imbalanced adjustment to n_R**

$$df = \frac{\left(\frac{\hat{\sigma}_T^2}{n_T} + \frac{\hat{\sigma}_R^2}{n_R}\right)^2}{\frac{\hat{\sigma}_T^4}{n_T^2(n_T - 1)} + \frac{\hat{\sigma}_R^4}{(n_R)^2(n_R - 1)}}$$

$$df^* = \frac{\left(\frac{\hat{\sigma}_T^2}{n_T} + \frac{\hat{\sigma}_R^2}{n_R^*}\right)^2}{\frac{\hat{\sigma}_T^4}{n_T^2(n_T - 1)} + \frac{\hat{\sigma}_R^4}{(n_R^*)^2(n_R - 1)}}$$

$$n_R^* = \min(n_R, 1.5n_T)$$

Dong, C., Weng, Y., and Tsong, Y., (2017), “Adjustment for unbalanced sample size for analytical biosimilar equivalence assessment”.

AMGEN

TIER 1 ISSUES

- **Power calculations assume**
  - **Reference product lots are independent**
  - **σ_R is known**
  - **Target "similarity criteria" $\mu_T - \mu_R = \frac{1}{8}\sigma_R$**
- **The patient's risk (type I error rate) increases due to**
  - **Uncertainty in the estimate of σ_R**
  - **Same data used to estimate EAC and confidence interval**
  - **Correlation of reference product lots**

AMGEN

OUTLINE

- Tier 1 Statistics
- **Tier 1 Example**
  - **EP2006 vs. US Neupogen®**

AMGEN

BLA 125553
EP2006, A PROPOSED BIOSIMILAR TO NEUPOGEN®

Figure 5. Content results of EP2006, US-licensed Neupogen and EU-approved Neupogen

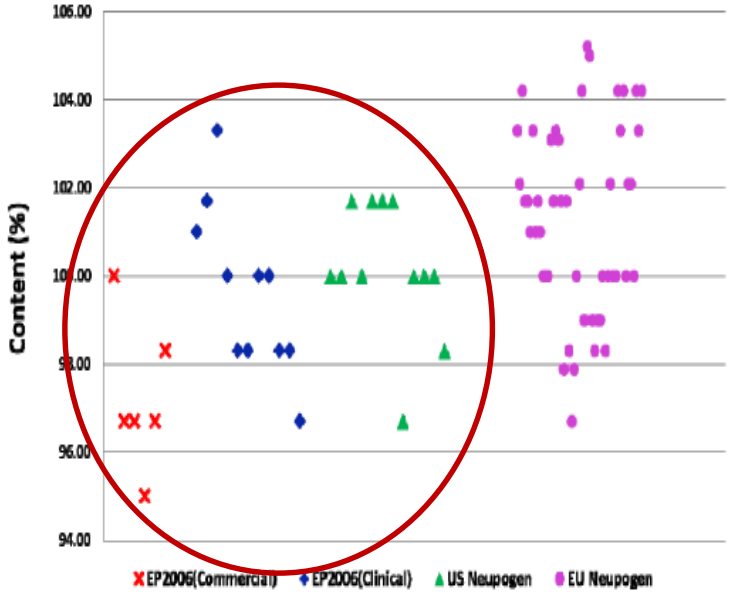


Table 4: Descriptive Statistics for Content (%)

Product	# of Batches	Min	Max	Mean	Standard Deviation	CV(%)
US-Neupogen	12	96.7	101.7	100.15	1.51	1.51%
EU-Neupogen	49	96.7	105.2	101.33	2.14	2.11%
Clinical EP2006	11	96.7	103.3	99.63	1.89	1.90%
Commercial EP2006	6	95.0	100.0	97.23	1.71	1.76%

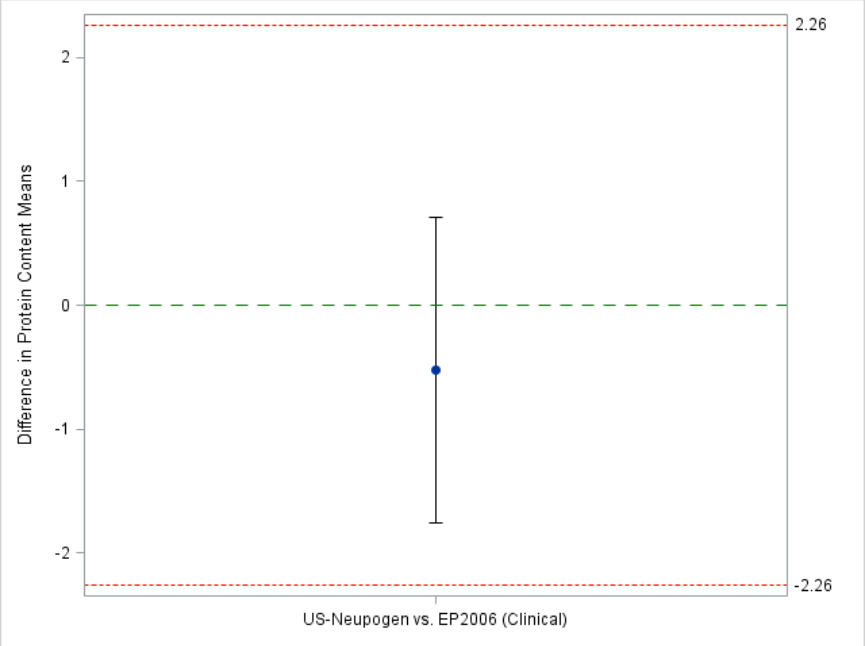
Table 5. Equivalence Testing Results for Content

Product	# batches	Comparator product	# batches	Statistical Equivalence?
EP2006 (Clinical)	11	US-Neupogen	12	**Yes**[a]
EP2006 (Commercial)	6	US-Neupogen	12	**No**[b]
EP2006 (Clinical)	11	EU-Neupogen	11	**Yes**[c]
EU-Neupogen	49	US-Neupogen	12	**Yes**[d]

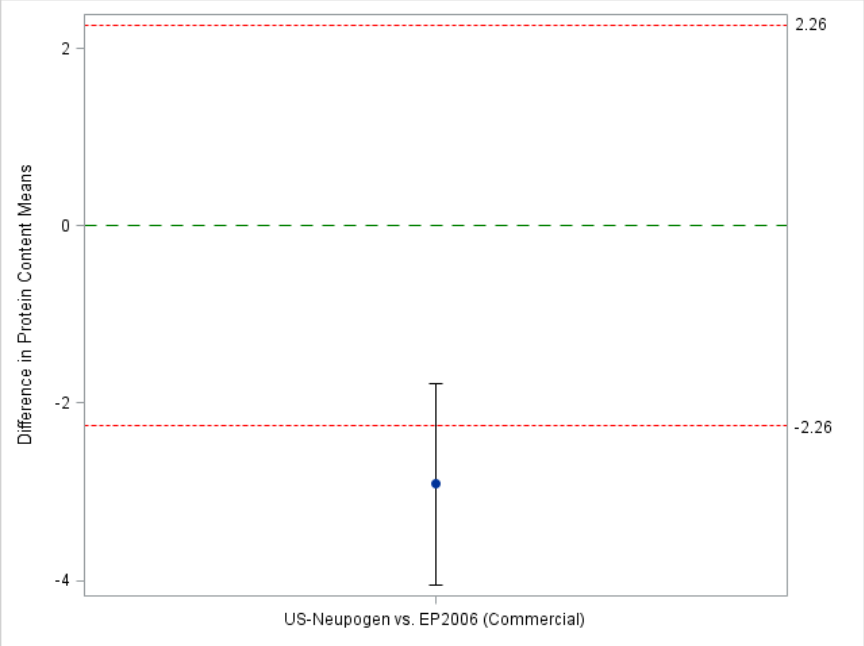
AMGEN

TIER 1 EQUIVALENCE TEST: US NEUPOGEN® VS. EP2006

US-Neupogen vs. EP2006 (Clinical)



US-Neupogen vs. EP2006 (Commercial)



The lower content of commercial EP2006 appears to be a manufacturing issue. However, this issue can likely be resolved by manufacturing and control strategies. Sandoz was asked to address this deficiency and provide data demonstrating that the proposed commercial EP2006 has the same "strength"[8] as US-licensed Neupogen.

OUTLINE

- Tier 1 Statistics
- Tier 1 Example
  - EP2006 vs. Neupogen®
- **Effect Size**

AMGEN

TIER 1: EFFECT SIZE

- **Burdick and Ramírez suggest testing using effect size**

$$H_0: \frac{\mu_T - \mu_R}{\sigma_R} \leq -1.5 \; or \; \frac{\mu_T - \mu_R}{\sigma_R} \geq 1.5$$

$$H_A: -1.5 \leq \frac{\mu_T - \mu_R}{\sigma_R} \leq 1.5$$

Glass' Effect Size

TIER 1: EFFECT SIZE

- ✓ EAC does not depend on σ_R
  - ✓ No double dipping
- ✓ σ_R is part of the test statistic
- ✓ Mitigates effect on patient's risk (type I error)

OUTLINE

- Tier 1 Statistics
- Tier 1 Example
  - EP2006 vs. Neupogen®
- Effect Size
- **Bayesian Tier 1 Statistics**

AMGEN

TIER 1: BEHRENS-FISHER PROBLEM

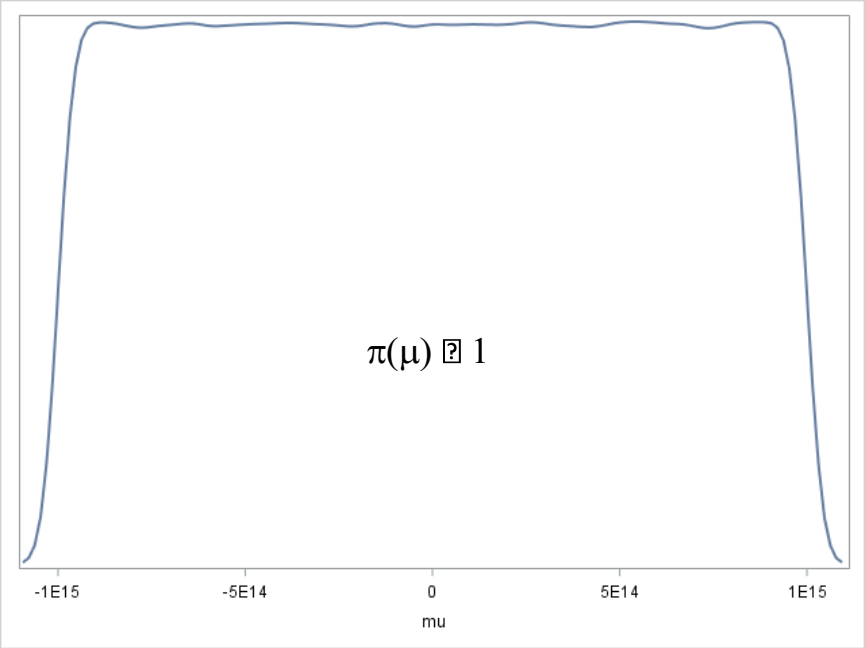
- t statistic = $\frac{\mu_T - \mu_R}{\sqrt{\sigma_T^2/n_T + \sigma_R^2/n_R}}$ = *signal-to-noise ratio*

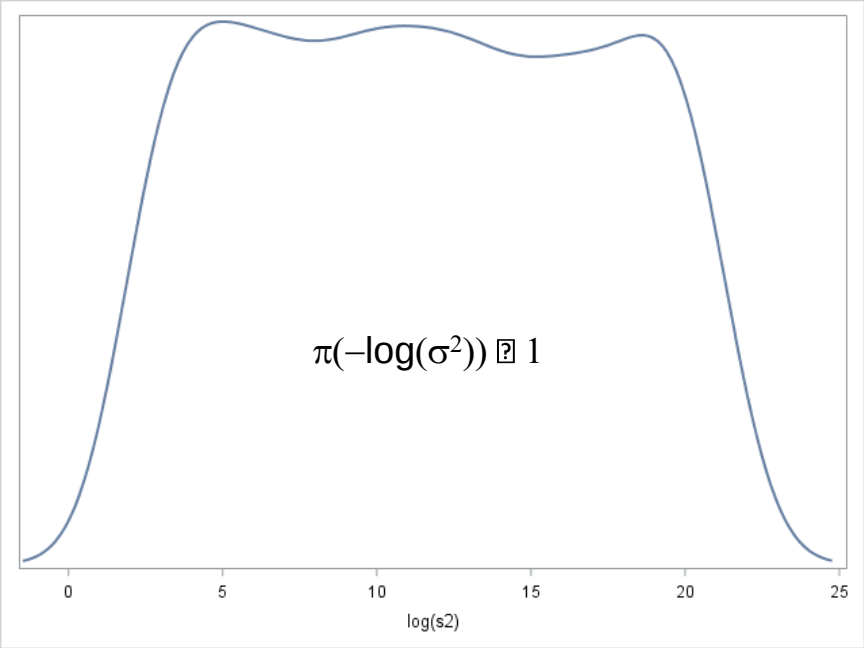
➢ **Distribution under H_0 not known**
  - **When variances not assumed equal**
- **What to do?**
  - **Satterthwaite degrees-of-freedom**
  - **Fisher's fiducial inference**
  - **Bayesian statistics**

TIER 1 BAYESIAN STATISTICS

- **Calculate the posterior distribution of** $\frac{\mu_T - \mu_R}{\sqrt{\sigma_T^2/n_T + \sigma_R^2/n_R}}$
- **Use diffuse priors on μ_R, μ_T, σ_R, and σ_T**
  - $\pi(\mu) \propto 1$ **uniform prior on [–∞, ∞]**
  - $\pi(-\log(\sigma^2)) \propto 1$ **uniform prior on log scale**
- **Compare 90% High Posterior Density interval to [-1.5, 1.5]**
- **Evaluate % of posterior density outside of [-1.5, 1.5]**

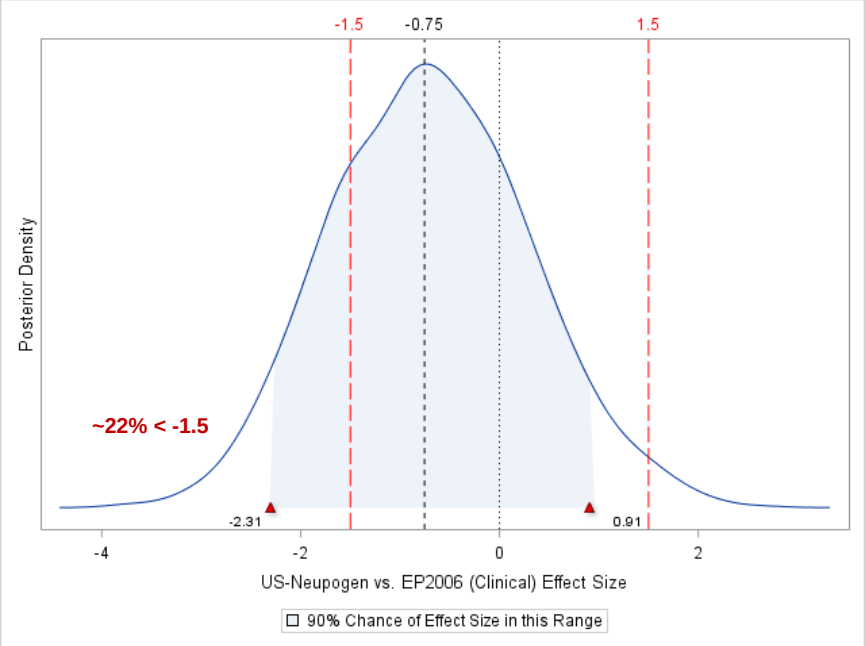
THE UNIFORM (FLAT) PRIORS



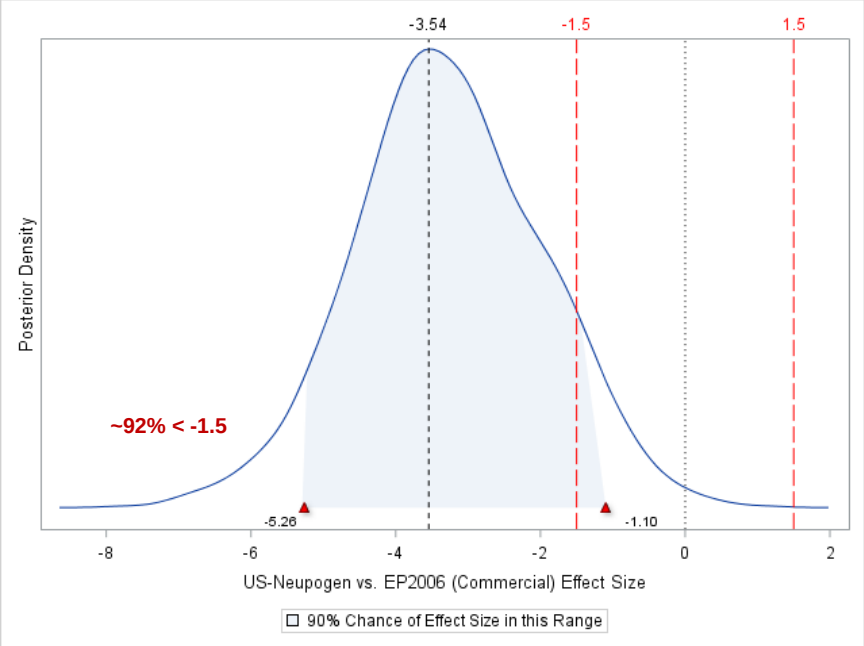


TIER 1: US NEUPOGEN® VS. EP2006
t STATISTIC POSTERIOR DENSITIES

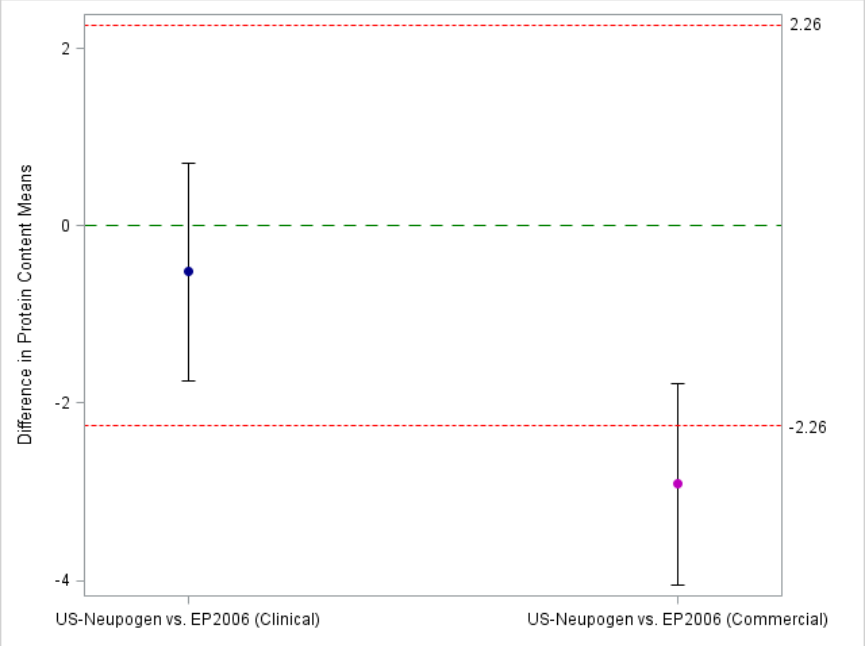
US-Neupogen vs. EP2006 (Clinical)

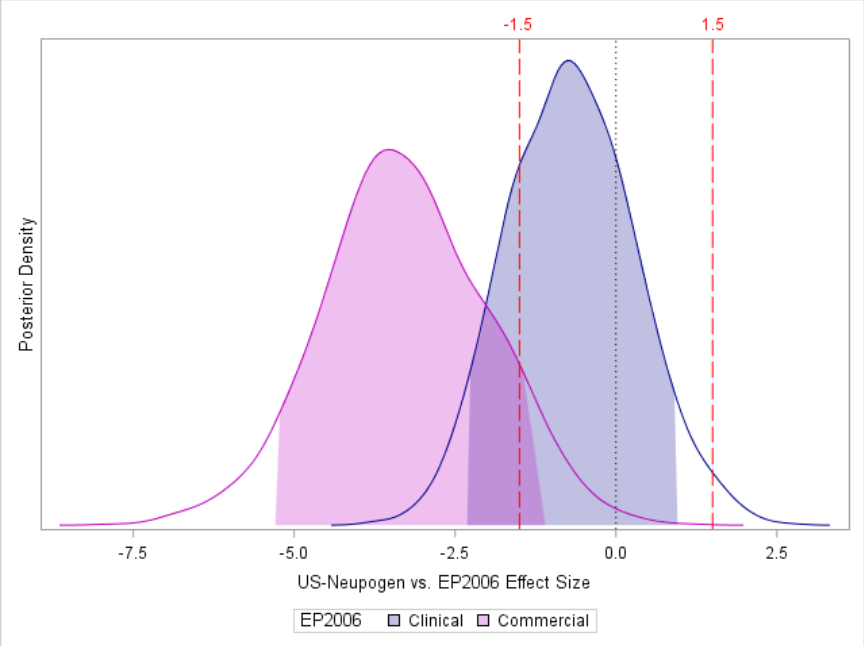


US-Neupogen vs. EP2006 (Commercial)



TIER 1: US NEUPOGEN® VS. EP2006 FREQUENTIST & BAYESIAN



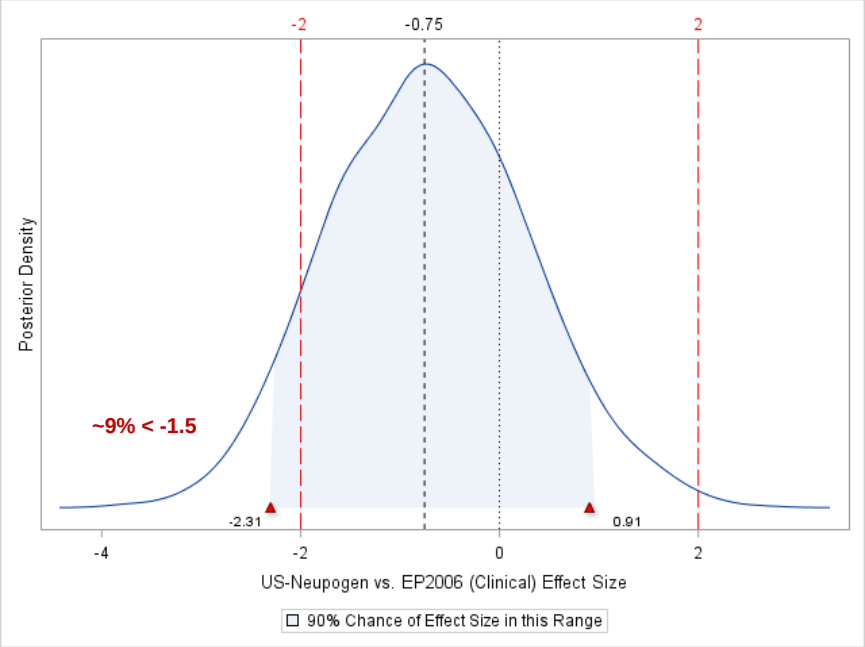


SUMMARY

- Evaluate *t* statistic (signal-to-noise ratio) directly
  - High posterior density (HPD) interval vs. ±EAC
  - % of posterior density outside ±EAC
- EAC defined as a multiple of the standard error of $\mu_T - \mu_R$
  - e.g., ±1.5
- No need to approximate degrees-of-freedom
  - No need for degrees-of-freedom
- Conclusions not based on large sample approximations

TIER 1: US NEUPOGEN® VS. EP2006
t STATISTIC POSTERIOR DENSITIES

US-Neupogen vs. EP2006 (Clinical)



US-Neupogen vs. EP2006 (Commercial)