



Equivalence Margin Considerations for Analytical Method Transfers

Daniel Obeng, Senior Biostatistician, Sanofi Specialty Care US

2017 IABS Conference

Agenda



- Testing for Mean Difference vs Equivalence
- Equivalence Margin and Acceptable Bias
- Power and Sample Size for TOST
- Considerations for the Equivalence Margin
- Considerations for the Acceptable Bias
- Thoughts from Industry
- Final Comments
- Acknowledgments

Testing for Mean Difference vs Equivalence

Test for Difference in Means

$$H_0: \mu_2 - \mu_1 = 0$$

$$H_1: \mu_2 - \mu_1 \neq 0$$

Objective: To prove that two population means are *statistically different* from one another

Test for Equivalence of Means

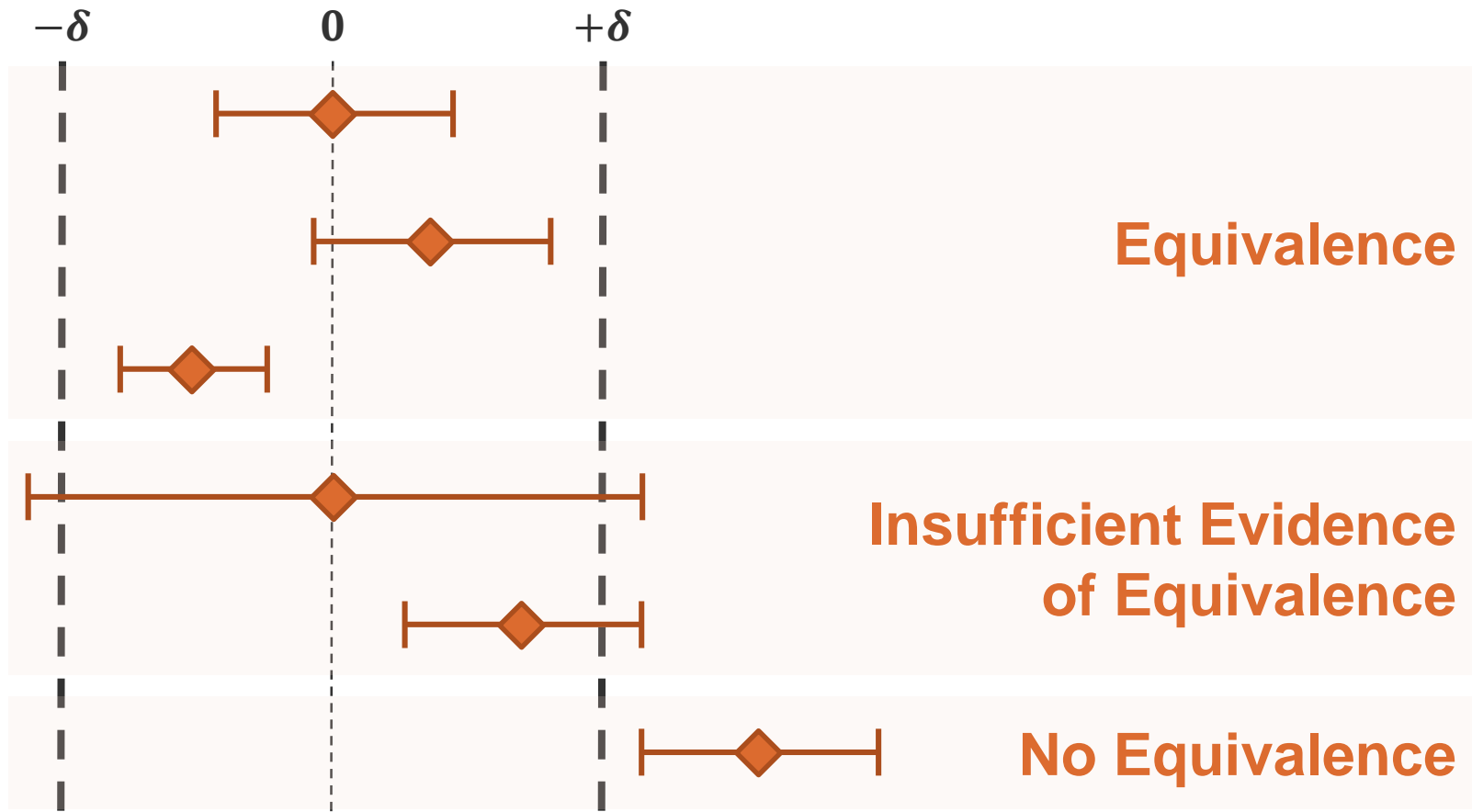
$$H_0: |\mu_2 - \mu_1| > \delta$$

$$H_1: |\mu_2 - \mu_1| \leq \delta$$

Objective: To prove that two population means are *statistically equivalent* to one another

- Each test is constructed to appropriately align the alternative hypothesis statement with power
- Equivalence is NOT the same as equality!
 - A data set could result in rejection of both null hypotheses
 - Selection of the equivalence margin, δ , is critical

Practical Equivalence Test



Equivalence Margin and Acceptable Bias

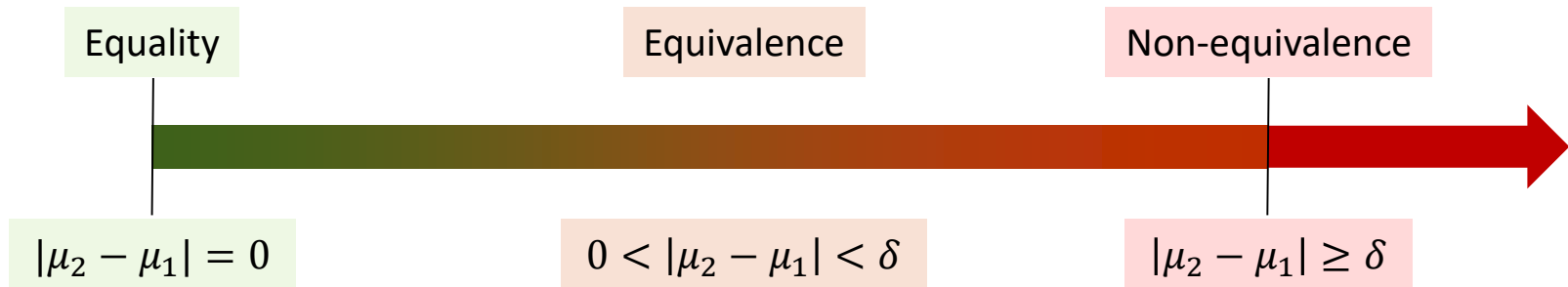
When thinking about practical equivalence, we have two expressions of the distance between the population means to consider...

- The **equivalence margin** is the largest difference between the population means that would still be considered a “comparable” or “tolerable” difference in means
- The **true** or **acceptable bias** is the true difference between the population means

Huh?!

Equivalence Margin and Acceptable Bias

Remember that **equivalence** is not the same as **equality**



- If the true difference between the population means is 0, then they can be considered **equal** (and by extension, equivalent)
- If the true difference between the population means is greater than 0 but less than the equivalence margin, then they may be considered **equivalent**
- If the true difference between the population means is greater than or equal to the equivalence margin, then they are **not equivalent**

Power and Sample Size for TOST

$$\text{Power} = \int_0^{\theta} \left[\Phi \left(\frac{\delta - \Delta}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} - t_{\alpha, \nu} \frac{x}{\sigma} \right) - \Phi \left(\frac{-\delta - \Delta}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} + t_{\alpha, \nu} \frac{x}{\sigma} \right) \right] \frac{1}{2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}} dx$$

Where

- n_1 and n_2 are the sample sizes for the two groups
- σ is the expected standard deviation
- α is the Type I error probability (significance level)
- $t_{\alpha, \nu}$ is a $100(1 - \alpha)^{\text{th}}$ quantile from a t -distribution with $\nu = n_1 + n_2 - 2$ degrees of freedom
- δ is the equivalence margin
- Δ is the true or acceptable bias, $|\mu_2 - \mu_1|$
- $\theta = \frac{\delta^2 \nu}{4 t_{\alpha, \nu}^2 \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$

Equivalence Margin and Acceptable Bias

We need to make decisions about both of these quantities

Equivalence Margin

What is the basis for determining the appropriate equivalence margin?

- **Process/science understanding?**
How much change is justified based on our understanding of the product/test method performance?
- **Process capability?**
How much change can our process tolerate and still meet specification consistently?
- **Desired statistical power?**
How wide does the equivalence margin need to be to deliver desired power for given n ?

Acceptable Bias

What is the appropriate assumption about the true bias between population means?

- **No bias is acceptable**
The two populations must, by nature or necessity, have equal means
- **Some bias is inevitable**
If the populations do not have to be equal, how much bias truly exists between the two population means? What is the basis for deriving this value?

Equivalence Margin for Test Method Transfers

Define δ based on science

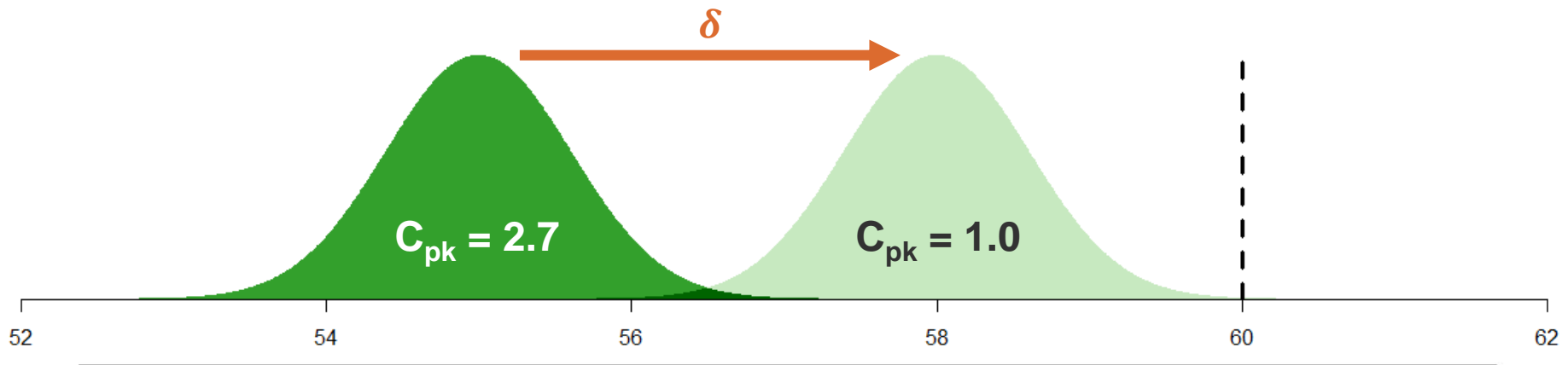
- This option is preferred by 99 out of 100 statisticians*
 - The difference between population means *considered [by the impacted SMEs] to be “comparable” or “tolerable”*
 - Defined based on prior understanding and experience, but independent of the data being immediately evaluated
 - Required sample size is driven by the desired power to prove equivalence
- This option is almost never used
 - SMEs typically don't have a framework for defining a limit of difference that can still be considered practically equivalent
 - Hesitance to draw a line in the sand for fear of unobtainable sample sizes
 - Culture has allowed this decision to migrate from practical determination to statistical determination “based on” historical data

*Estimated from IABS coffee break conversations using the a-mathematical technique of qualitative guestimation

Equivalence Margin for Test Method Transfers

Define δ based on process capability

- This option seems like a reasonable blend of practical and statistical considerations
 - The “science” provides the quality limit(s) (e.g., specification limits)
 - The “statistics” provide an understanding of risk to fail the quality limit(s) (e.g., C_{pk})
 - The equivalence margin is the distance that the mean can shift while maintaining a tolerable probability of meeting the quality limits



Equivalence Margin for Test Method Transfers

Define δ based on process capability

- This option requires a specific understanding of equivalence
 - Based on the initial capability of the process to meet specification, the distance between the two populations may be *quite* different but still be considered practically equivalent
 - Use of process-based capability to evaluate acceptable shift in analytical method may be too far of a reach (i.e., how much shift could/should be attributable to analytical method?)
- This option has important estimation considerations and limitations
 - Need to account for estimate uncertainty using confidence limits?
 - If no quality limits exist, process capability cannot be calculated and this approach is not applicable

Equivalence Margin for Test Method Transfers

Define δ based on statistical power requirements

This option leans most heavily on statistics

Step 1: Identify the sample size that can be most reasonably obtained for analysis

Step 2: Determine the desired power for the equivalence test

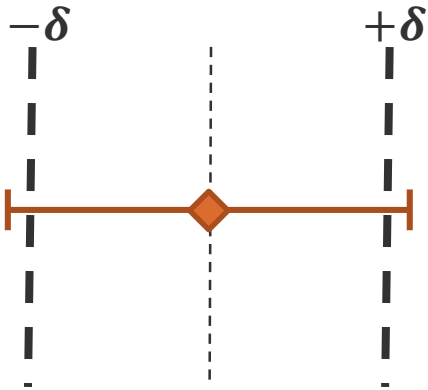
Step 3: Define the acceptable bias

Step 4: Identify the equivalence margin δ that yields the specified power for the given sample size

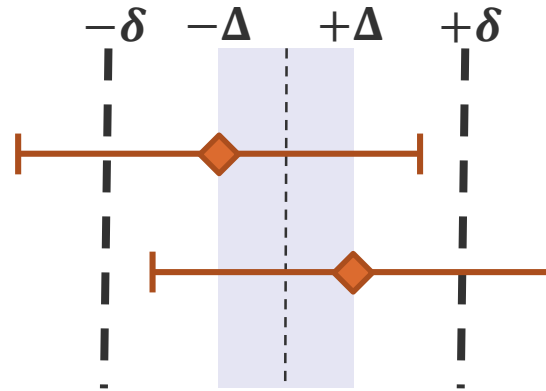
- Pros: Provides an equivalence margin that will yield the desired size and power for the amount of data available
- Cons: The equivalence margin doesn't mean anything and can be rather large for smaller sample sizes

Equivalence Margin for Test Method Transfers

$$\text{Let } \Delta = |\mu_2 - \mu_1|$$



Power is the probability that the confidence interval is fully contained within the equivalence margin assuming the true means are equal



Power is the probability that the confidence interval is fully contained within the equivalence margin assuming the true means differ by Δ

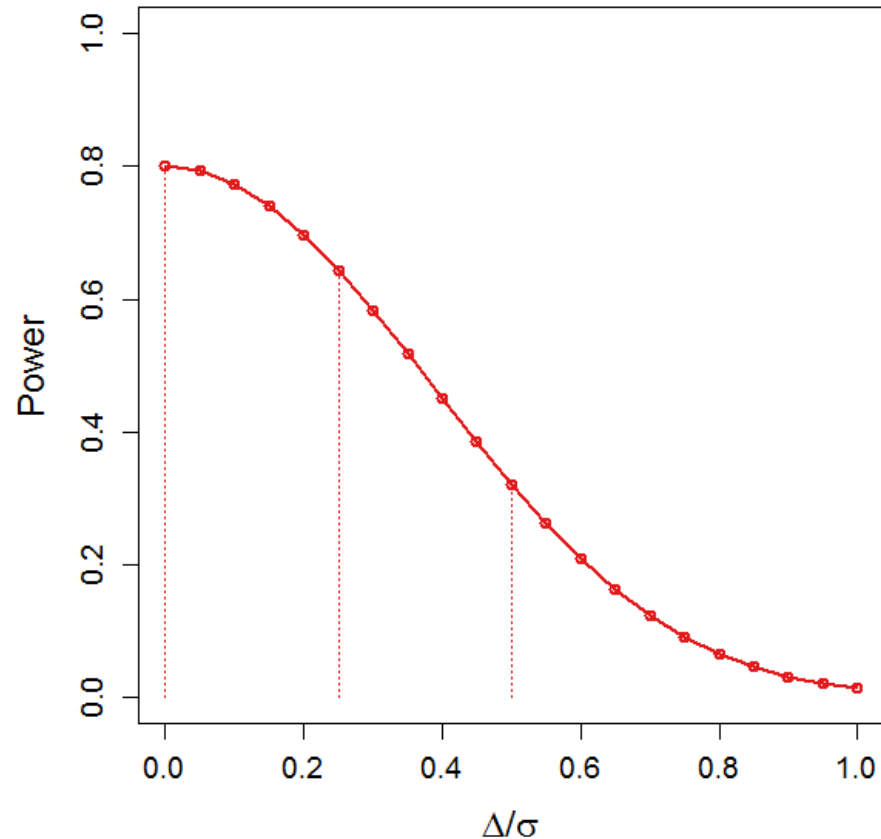
Equivalence Margin for Test Method Transfers

Cost of assuming no acceptable bias?

The equivalence margin for 80% power when $n_1 = n_2 = 25$ and $\Delta = 0\sigma$ is $\delta = \mathbf{0.840\sigma}$.

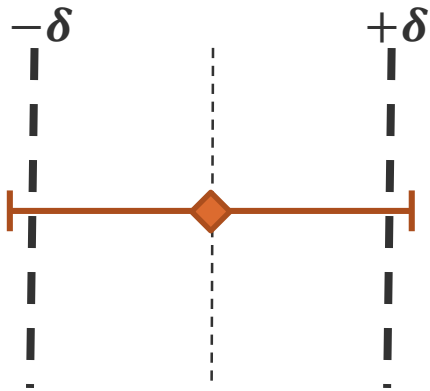
If the true bias is...

- $\Delta = 0.25\sigma$, then the power drops to approximately **64%**
- $\Delta = 0.5\sigma$, then the power drops to approximately **32%**

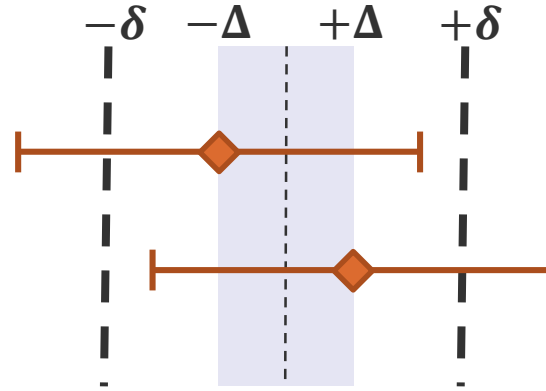


Equivalence Margin for Test Method Transfers

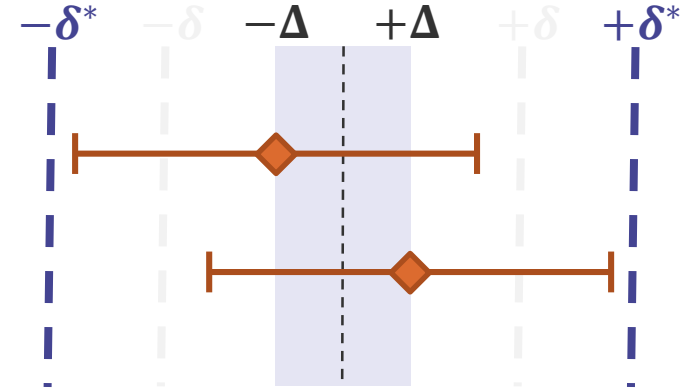
Let $\Delta = |\mu_2 - \mu_1|$



Power is the probability that the confidence interval is fully contained within the equivalence margin assuming the true means are equal

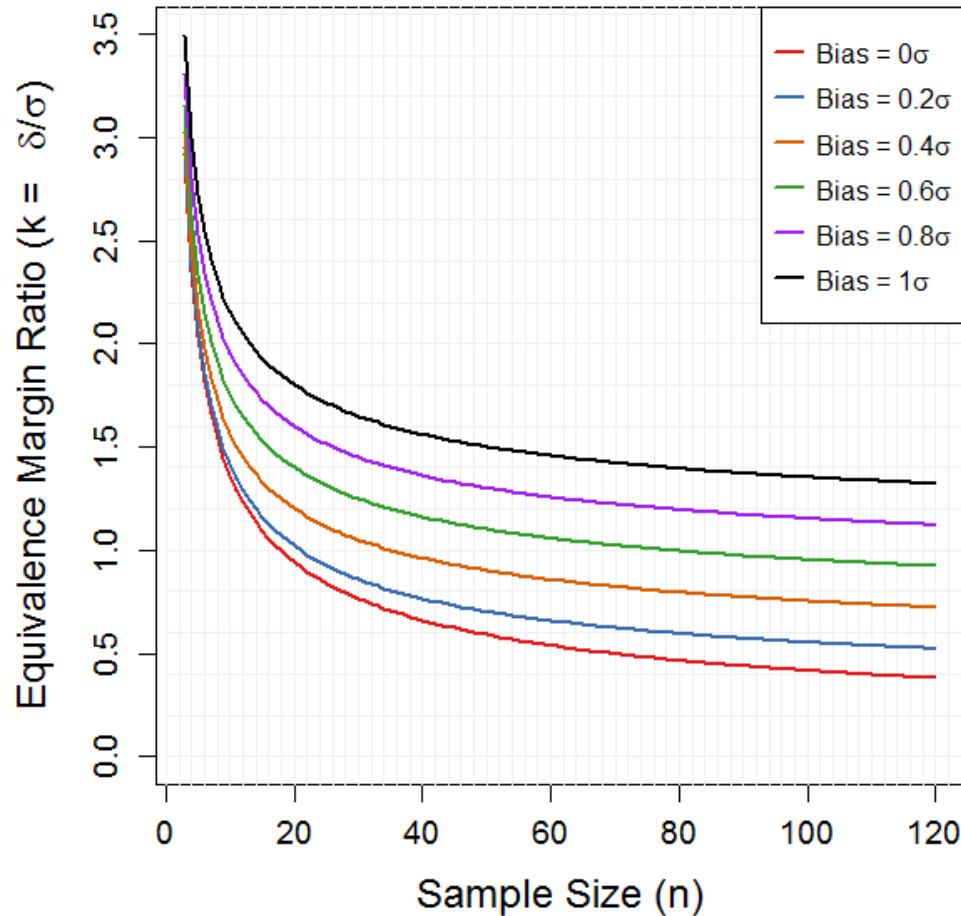


Power is the probability that the confidence interval is fully contained within the equivalence margin assuming the true means differ by Δ



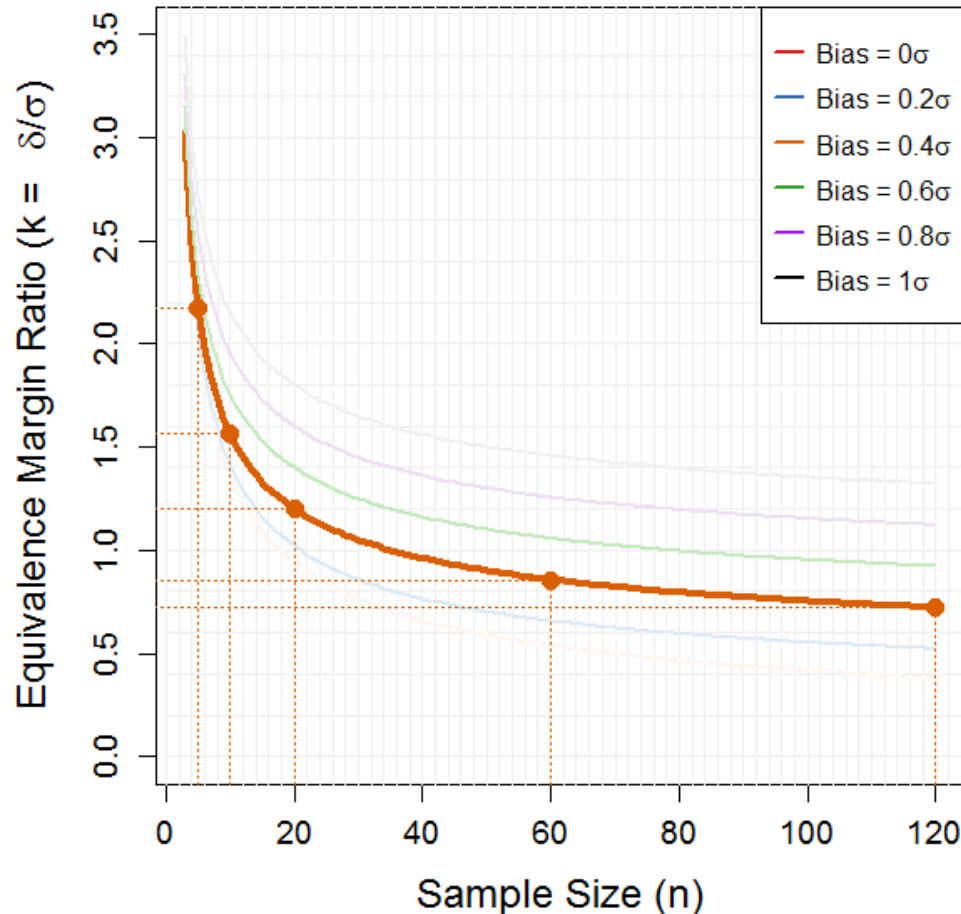
Adjust the equivalence margins to account for acceptable bias (and sample size)

Equivalence Margin for Test Method Transfers



For a given acceptable bias, any equivalence margin value on the curve will have 80% power to prove equivalence at the selected sample size (95% confidence level)

Equivalence Margin for Test Method Transfers



For a given acceptable bias, any equivalence margin value on the curve will have 80% power to prove equivalence at the selected sample size (95% confidence level)

Example:

To achieve a power of 80% for an acceptable bias of 0.4 σ ...

- if $n = 5$, set $\delta = 2.176\sigma$
- if $n = 10$, set $\delta = 1.565\sigma$
- if $n = 20$, set $\delta = 1.201\sigma$
- if $n = 60$, set $\delta = 0.857\sigma$
- if $n = 120$, set $\delta = 0.722\sigma$

Acceptable Bias for Test Method Transfers

Is the acceptable bias 0 or not 0?

$$\Delta = 0$$

- Most conservative approach, providing greatest “protection” of patient
- “No expert in biological sciences can accept $\delta > \sigma_S$ used for the method transfer studies...” (Shen & Xu, 2017)
 - For 95% confidence, 85% power, $n=20$, $\delta = 0.99\sigma_S$ ($0.94\sigma_S$ for 80% power)
- Difficult to justify basis for bias based on “historical” data

$$\Delta \geq 0$$

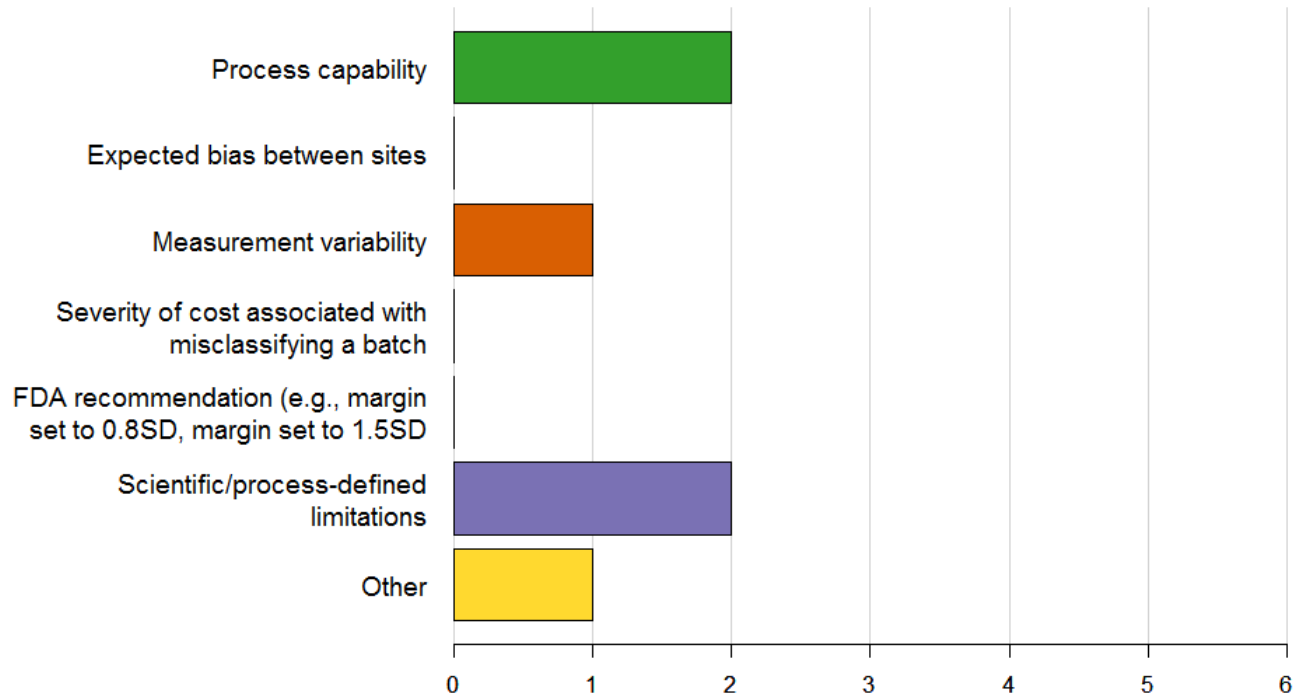
- In many cases, it’s unrealistic to assume no bias when transferring from SU to RU
- Acceptable bias may be derived from accuracy criterion (i.e., amount of mean shift that would allow RU method to still be deemed suitably accurate)
- Industry discussion/best practices needed here

Thoughts from Industry

Surveyed 6 industry statisticians from different companies about analytical test method transfers

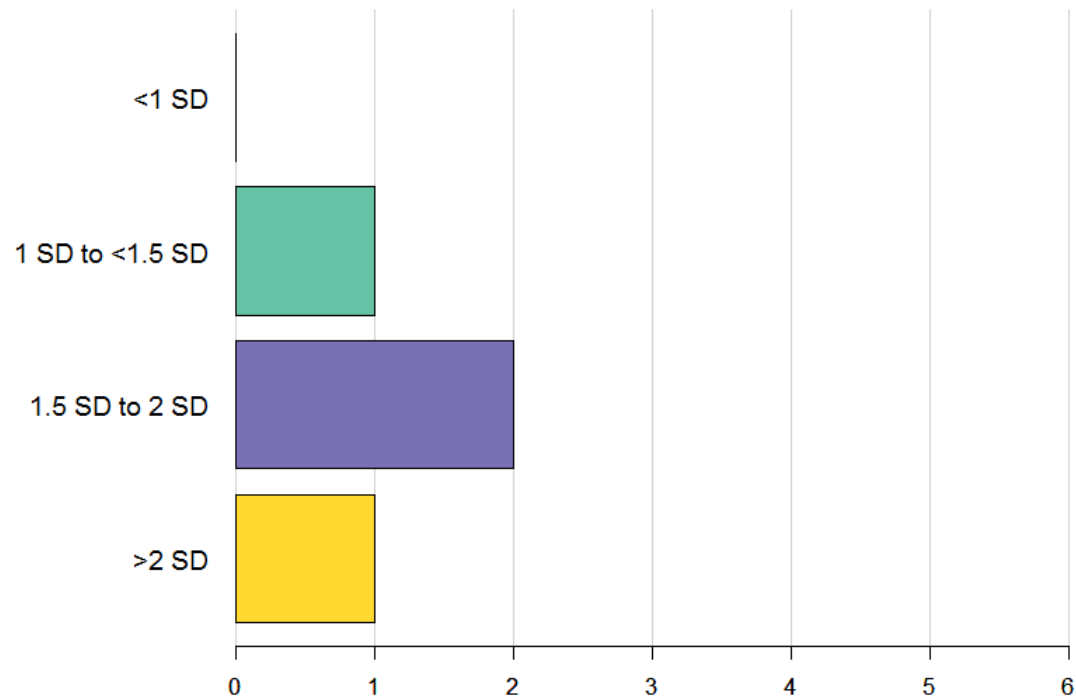
Here's what they had to say...

In your CURRENT practice, which of the following concepts most often informs your equivalence margin determinations?

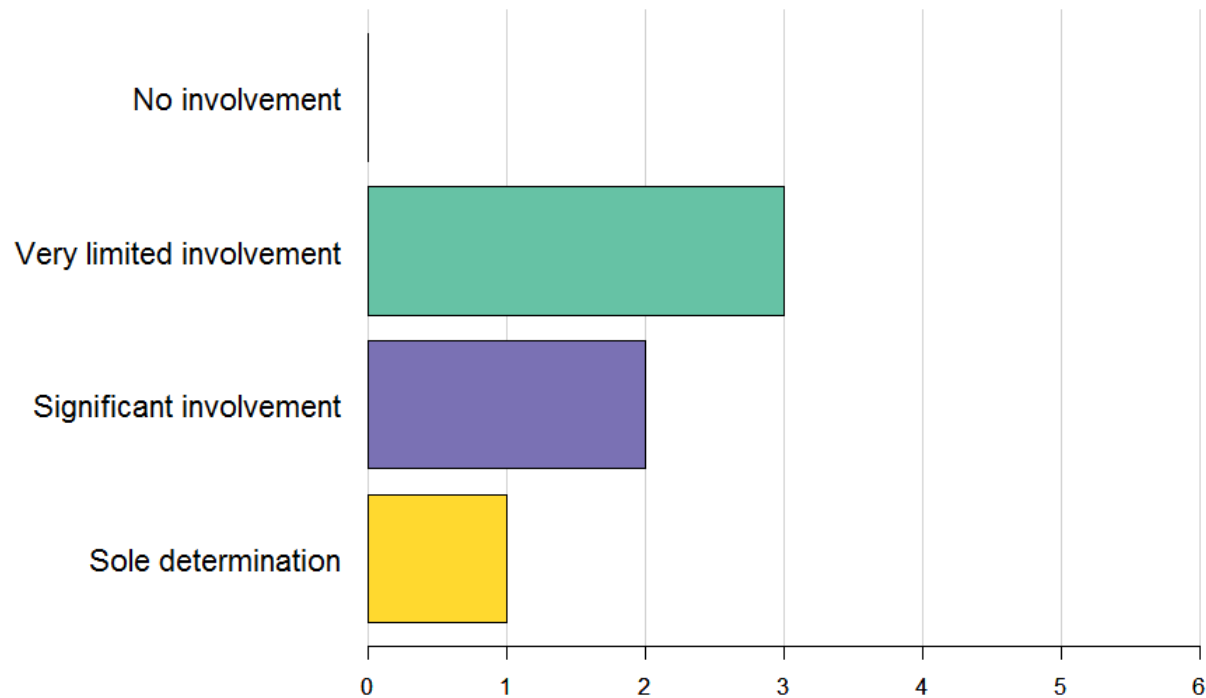


Other: Process capability, expected bias, historical variability, feasible sample sizes

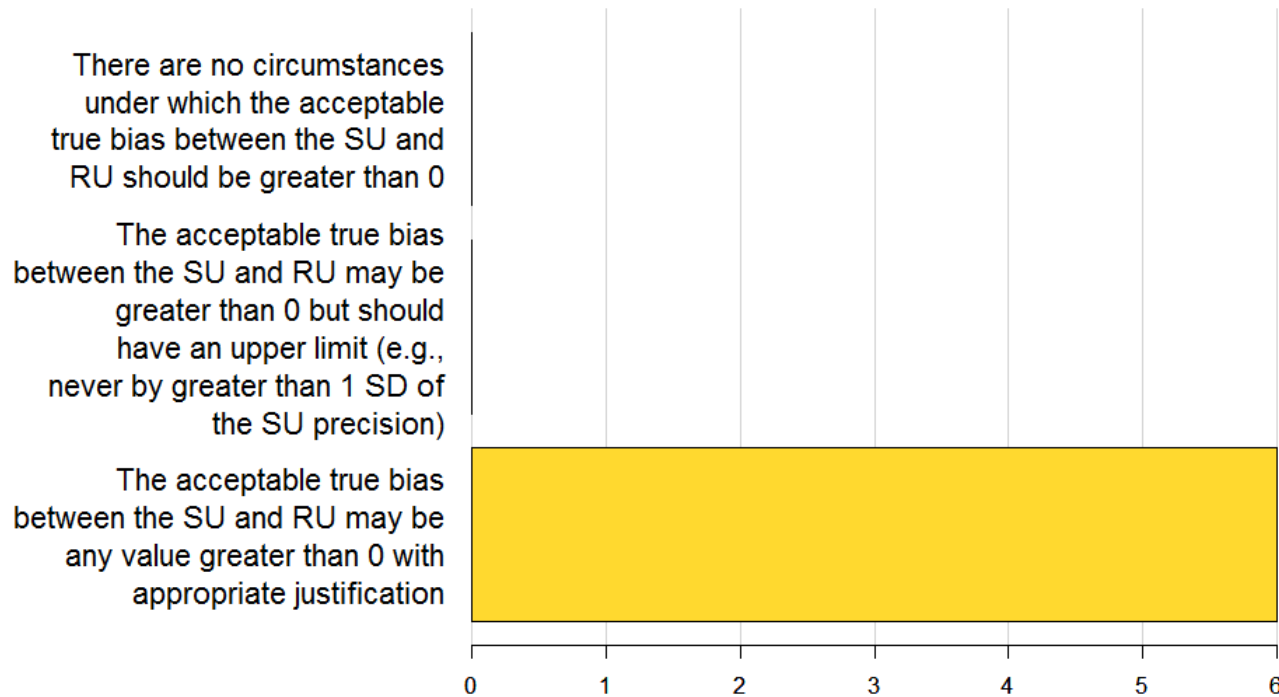
Select the range of typical equivalence margins you use for analytical method transfers in units of the sending unit standard deviation



To what extent are non-Statistician SMEs involved in the determination of the analytical method transfer equivalence margin?



Which of the following statements best reflects your view regarding the acceptable true bias between the Sending Unit (SU) and Receiving Unit (RU)?



Which of the following statements best reflects your view regarding the acceptable true bias between the Sending Unit (SU) and Receiving Unit (RU)?

Comments

When the receiving unit's precision compensates for the bias. Conservation of total error.

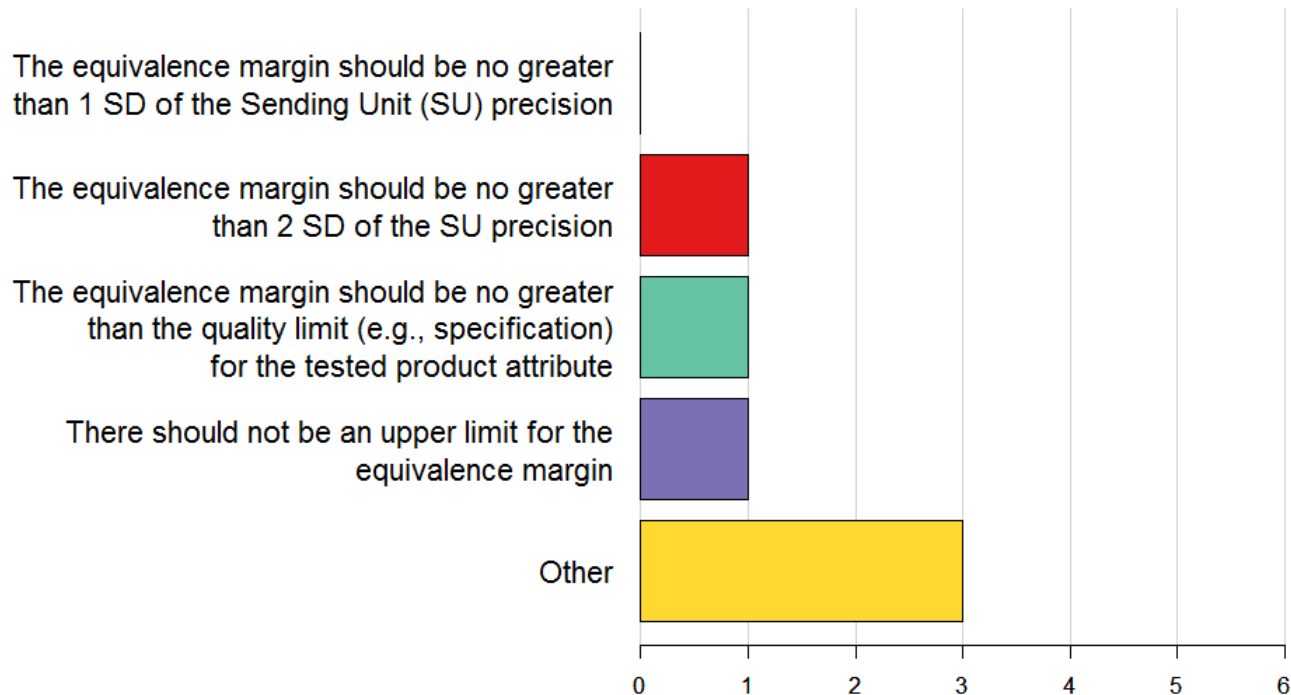
The determination of acceptable bias should take into account specification limits/target limits to determine potential OOS rates at the receiving lab post transfer - producer's risk. This is then more a business decision as the OOS rates affect repeats etc.

It depends on the amount of room you have between the specs and the RU mean

It must be demonstrated or justified that the acceptable true bias has no impact on the analytical method's fitness for use in determining "good" product quality. This means that theoretically this bias could be as great as 4 or 5 SD, but if results are still within specification with high capability then it is acceptable. Note that this is not a widely held view, and in practice specifications are not usually wide enough anyway to allow for this possibility, but the idea is presented here for the sake of argument

The equivalence margin should be based on multiple factors: safety, efficacy, risk of out-of-specification, etc. A multiple of method variability may not be meaningful. In my opinion, whether the eq. margin should be 1 or 2 SD is moot.

In your opinion, what is the appropriate upper limit for an equivalence margin?



Bounded by consideration of process capability in the receiving unit.

Other:

Should be based on the corresponding type 1 and type 2 risks for given transfer design (sample sizes)

The amount of shift that keeps OOS at an acceptable level. Somewhat similar to <1010> approach.

Final Thoughts

- The use of the TOST for analytical test method transfer is fairly universally accepted
- There remains a need for practical discussion on best practices for determination of the equivalence margin
- Consideration for TOST variants
 - Equivalence for paired data
 - Equivalence of process variances

Acknowledgments

- Fei Wang
- Michael Rose
- Xian-Jie Yu
- Todd Regh

- Survey respondents



THANK YOU!