Demonstration of Comparability in Accelerated or Stressed Stability Studies Using Graphical and Statistical Methods

Richard K. Burdick
CMC Strategy Forum
January 27, 2020

Introduction

- Regulatory bodies recognize and accept change as a normal part of manufacturing in a cGMP environment.
- To assess the impact of process changes, stability data (e.g., measurements of drug product over time) at recommended storage conditions are required.
- The degradation rate (slope over time) of a post-change process is compared to the pre-change process in order to determine if the slopes are comparable.
- Since slopes at the recommended storage conditions are often quite small, product is also exposed to non-recommended storage conditions (e.g., higher temperature) in order to quickly accelerate degradation.

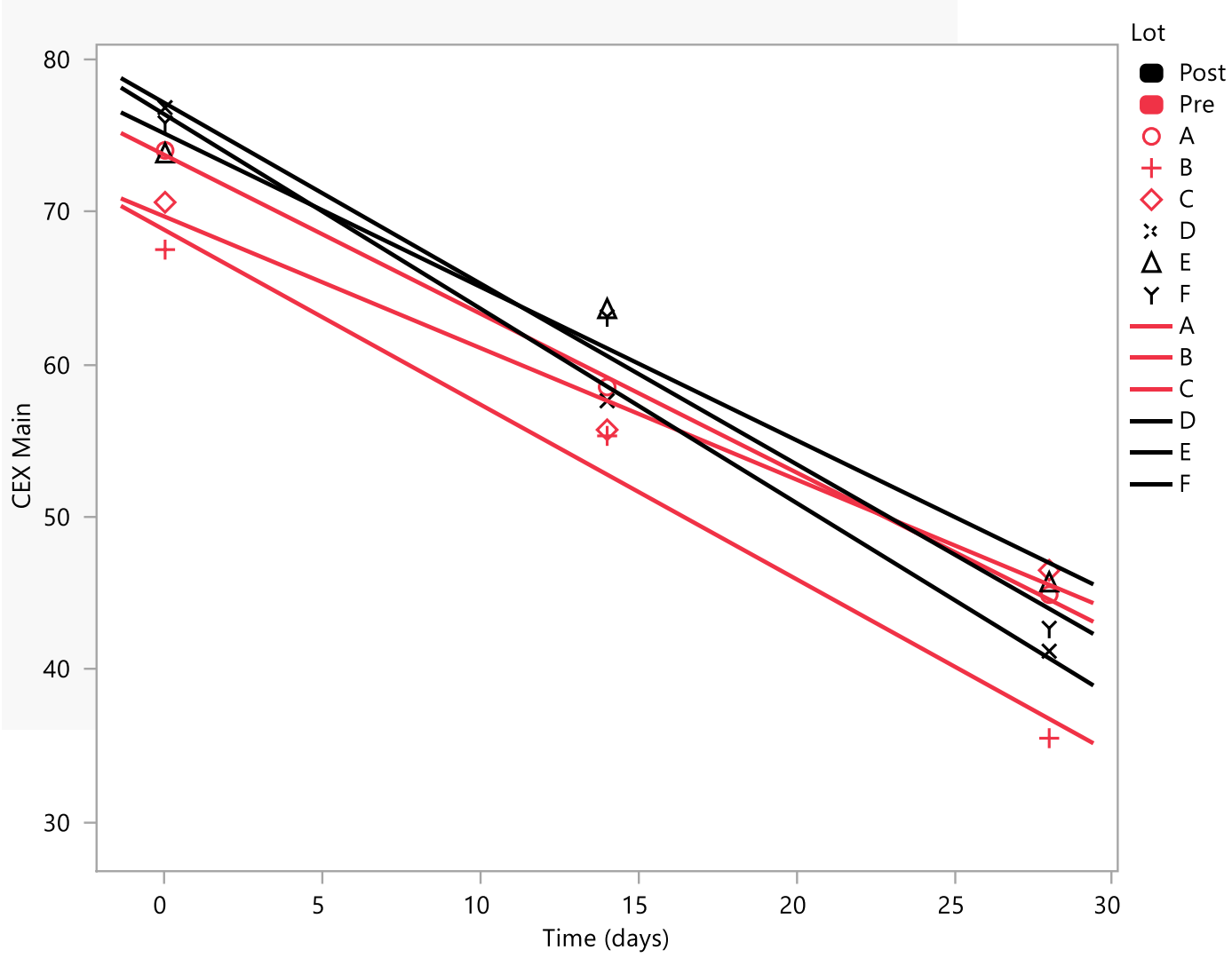
Example data set

- 3 pre-change lots
- 3 post-change lots
- Response is % relative main peak for CEX analytical method
- Time periods are 0, 14, and 28 days at non-recommended temperature conditions.

Aways, Always, Always Rule (L. Hare)

- Plot the data
- Corollary
  - Look at it!

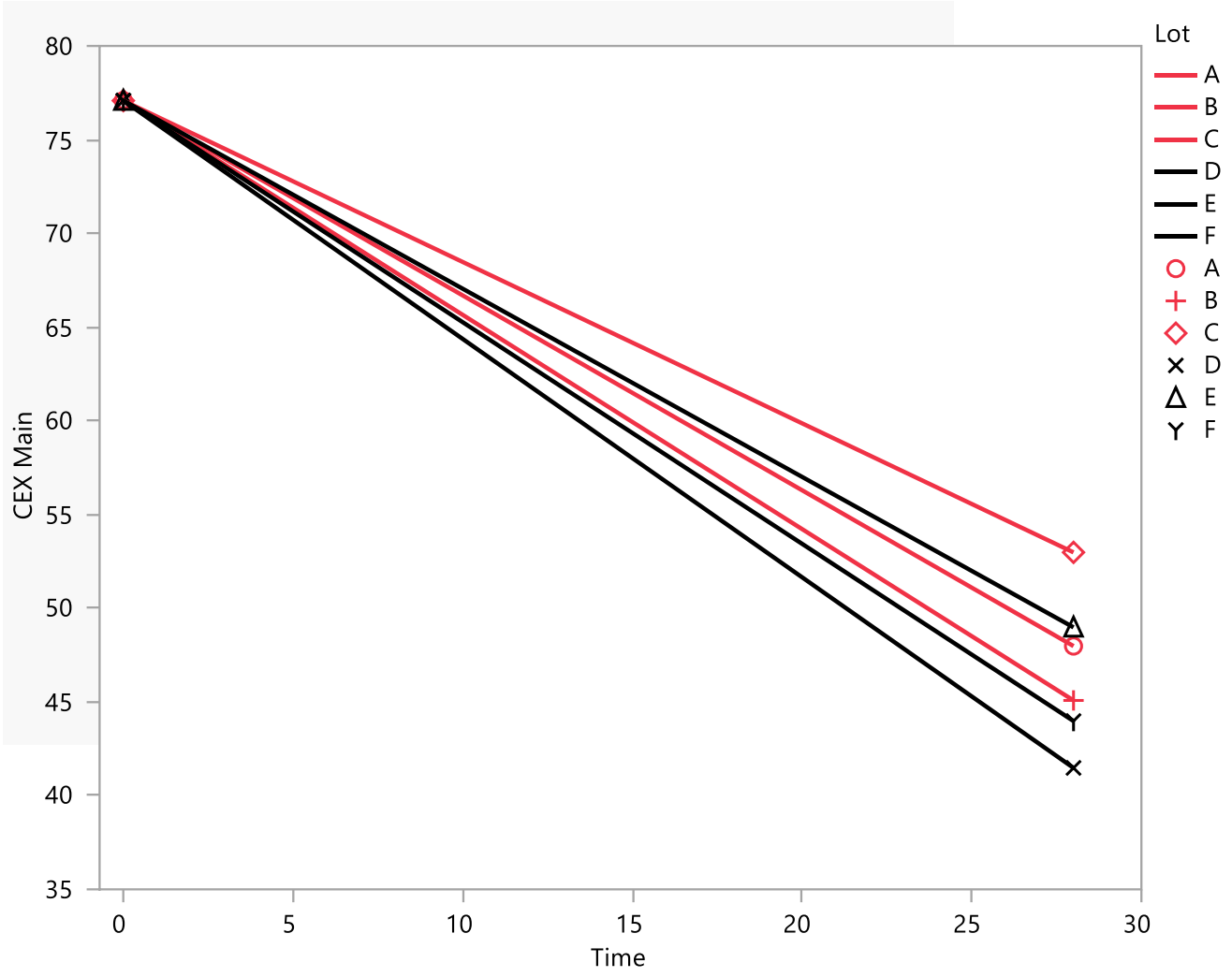
Stability plots by lot and group



Estimated slopes and y-intercepts

Group	Lot	Y intercept	Slope
Pre	A	73.7	-1.04
Pre	B	68.8	-1.14
Pre	C	69.7	-0.86
Post	D	76.3	-1.27
Post	E	75.1	-1.00
Post	F	77.1	-1.18

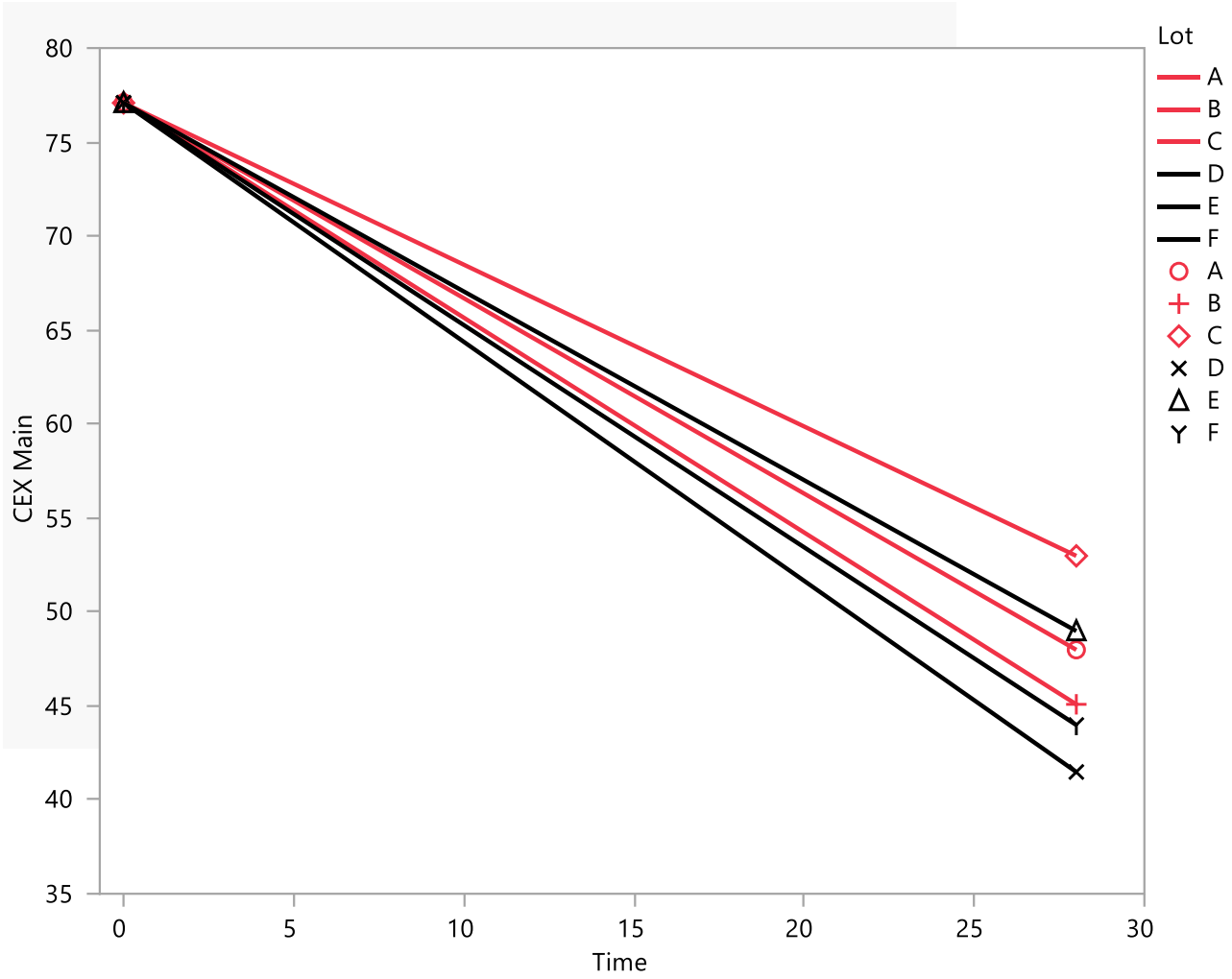
Stability plots at same y-intercept



Conversion can be done in Excel

Lot	Time	Predicted	Adjusted	Slope
A	0	73.683333	77.08333	
A	28	44.583325	47.98	-1.04
B	0	68.766667	77.08333	
B	28	36.766671	45.08	-1.14
C	0	69.65	77.08333	
C	28	45.550008	52.98	-0.86
D	0	76.333333	77.08333	
D	28	40.733321	41.48	-1.27
E	0	75.083333	77.08333	
E	28	46.983345	48.98	-1.00
F	0	77.083333	77.08333	
F	28	43.983329	43.98	-1.18
	Max	77.083333		

What can we conclude?



A convenient metric to describe differences is the **effect size (ES)**

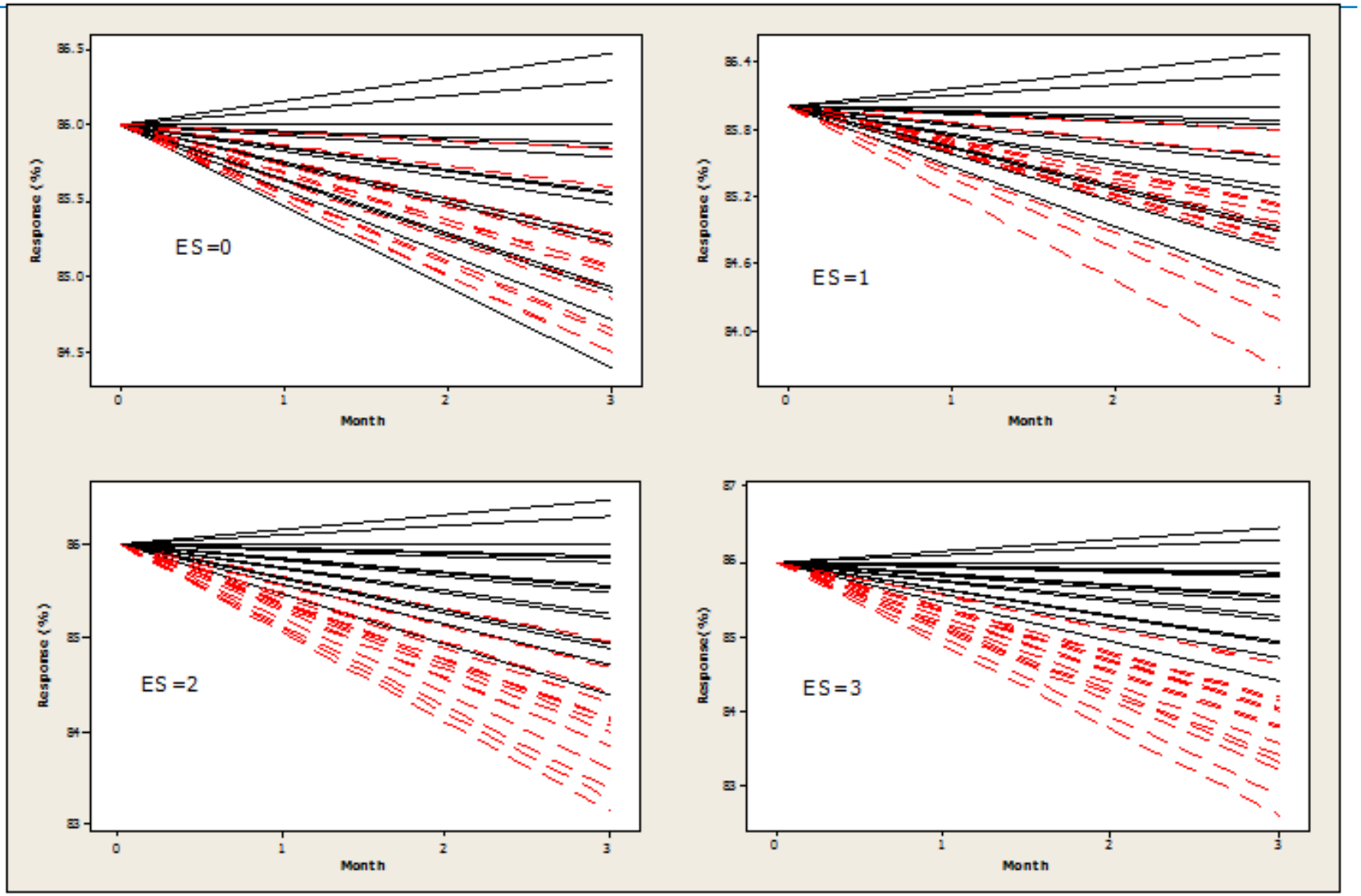
What is an effect size (ES)?

- Effect size (ES) is the scaled difference in process slopes.
- In particular,

$$ES = \frac{\text{Difference between pre- and post-change slopes}}{\text{Standard deviation of pre-change slopes}}$$

- In this example, it is the difference in the average of 3 pre-change slopes and 3 post-change slopes, divided by the standard deviation of the pre-change slopes.

Visualization of effect sizes (ES) for differences of slopes

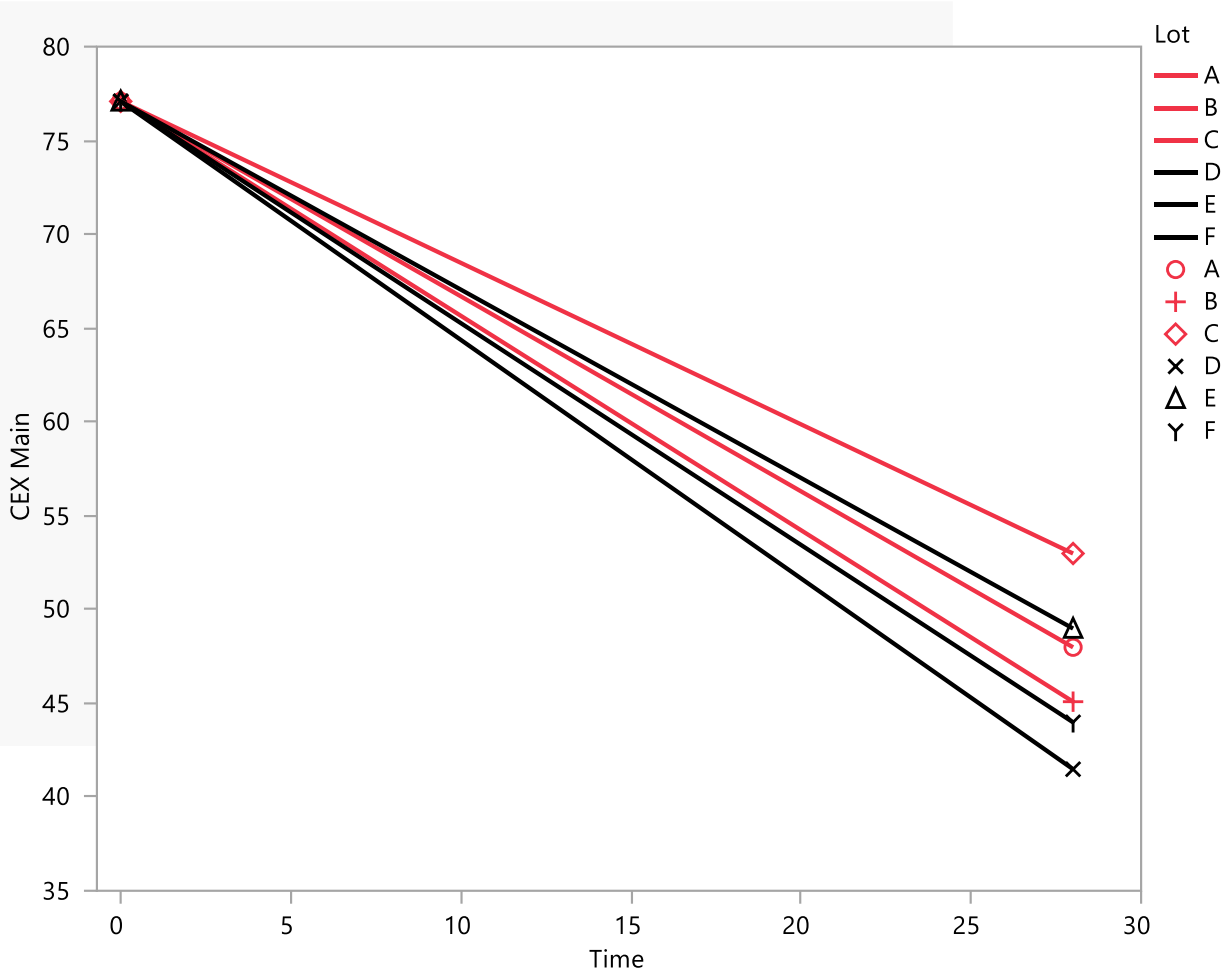


Subject matter expert should define comparability in terms of ES

Overlap as function of effect size (Lei and Olson, 2010)

Effect Size	Area overlapped by two population distributions
0	100%
1	62%
2	32%
3	13%

What can we conclude in our example?



Here we have 1 of 3 (33%) post-change overlap pre-change range. (Lot E)

Possible decision rules

1. All post-change lots (black) fall within the min-max slope range of pre-change lots (red).
2. At least one post-change lot fall within the min-max slope range of the pre-change lots.
3. All post-change slopes fall in a quality range based on the pre-change slopes.
   - The quality range approach has been recommended by FDA for demonstrating analytical similarity of product *means* (2019).
   - This same approach can be used for product *slopes*.

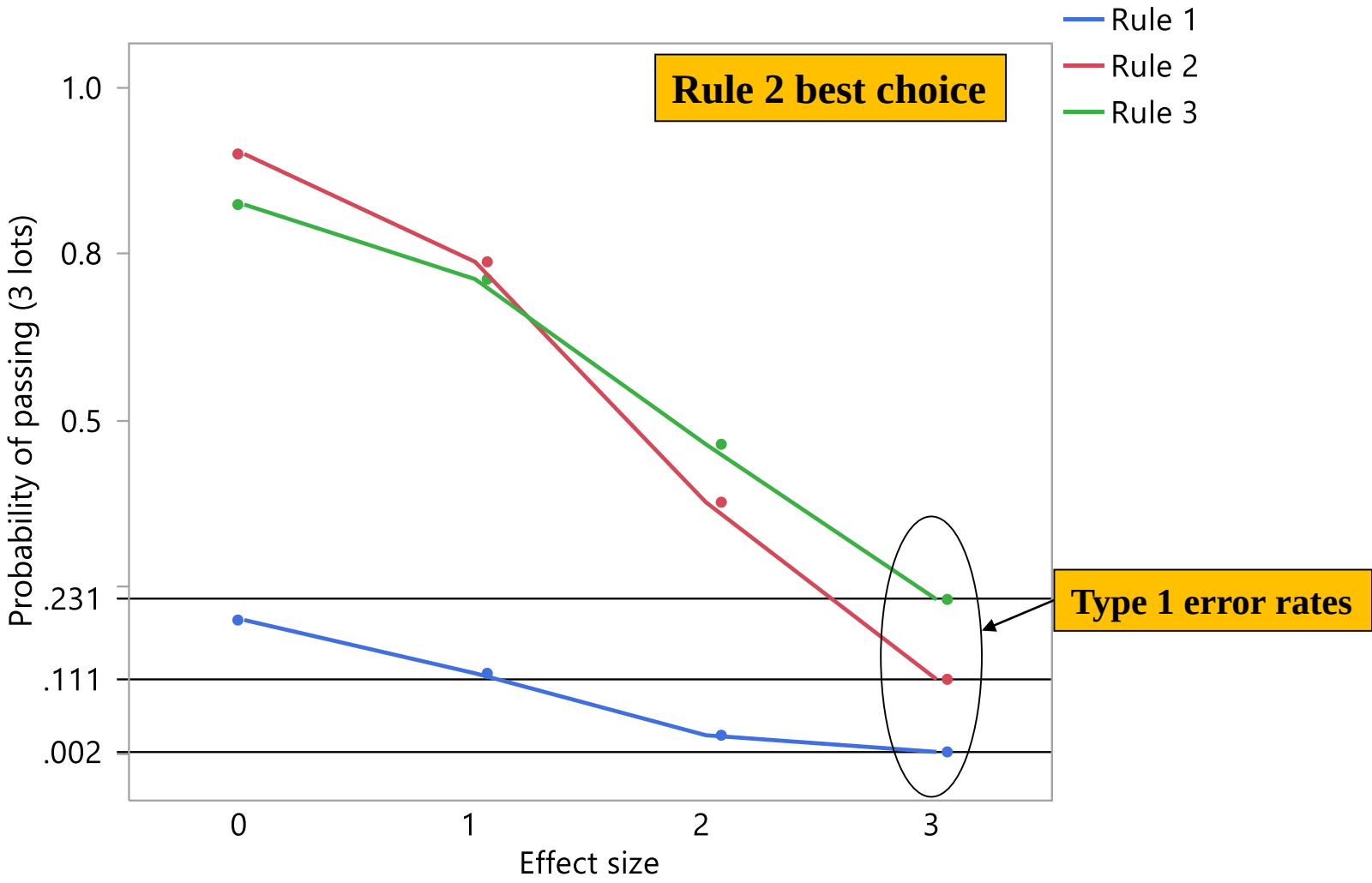
$$LQR = \bar{b}_{Pre} - K \times \text{std dev } b_{Pre}$$

$$UQR = \bar{b}_{Pre} + K \times \text{std dev } b_{Pre}$$

How do we select an appropriate rule?

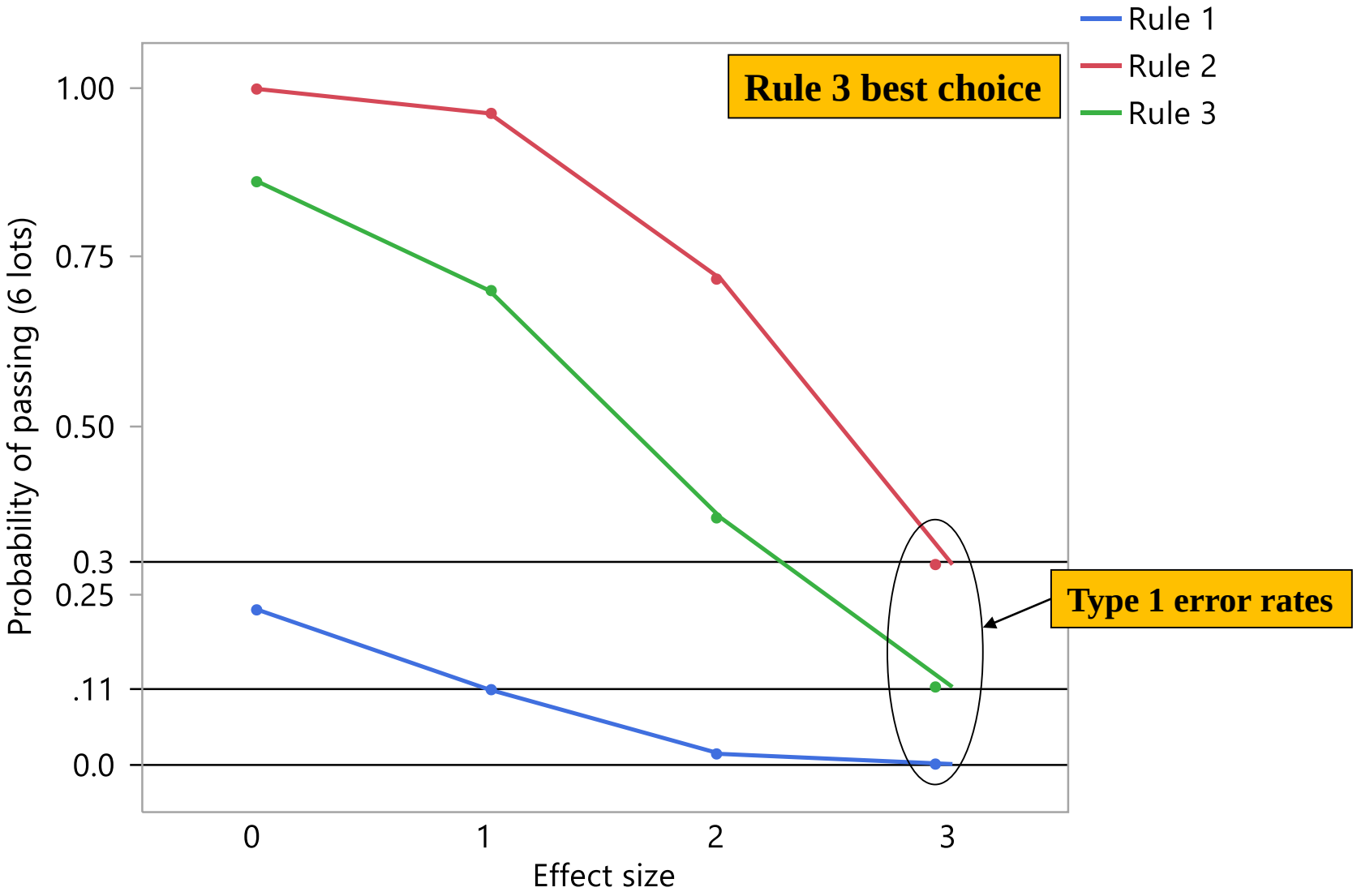
- Consider decision errors that can be made and then determine the desired probability of error for each rule.
  - Statisticians call these power calculations.
  - Type 1 error: Claiming comparability when such is not the case.
  - Type 2 error: Failing to claim comparability when processes are comparable.
- What does it mean to be comparable?
  - Subject matter experts can define comparability in terms of an acceptable ES value.
  - For today, we will use ES=3 to define comparability.
  - Regulators are interested in this definition, and want sponsors to answer the question,
    - **"How bad do things need to be before you will conclude the processes are not comparable?"**

Power curves with 3 lots with comparability defined as ES=3



We want a high probability of passing for ES=0 and low probability for ES=3 (which is the type 1 error rate).

Power curves with 6 lots with comparability defined as ES=3

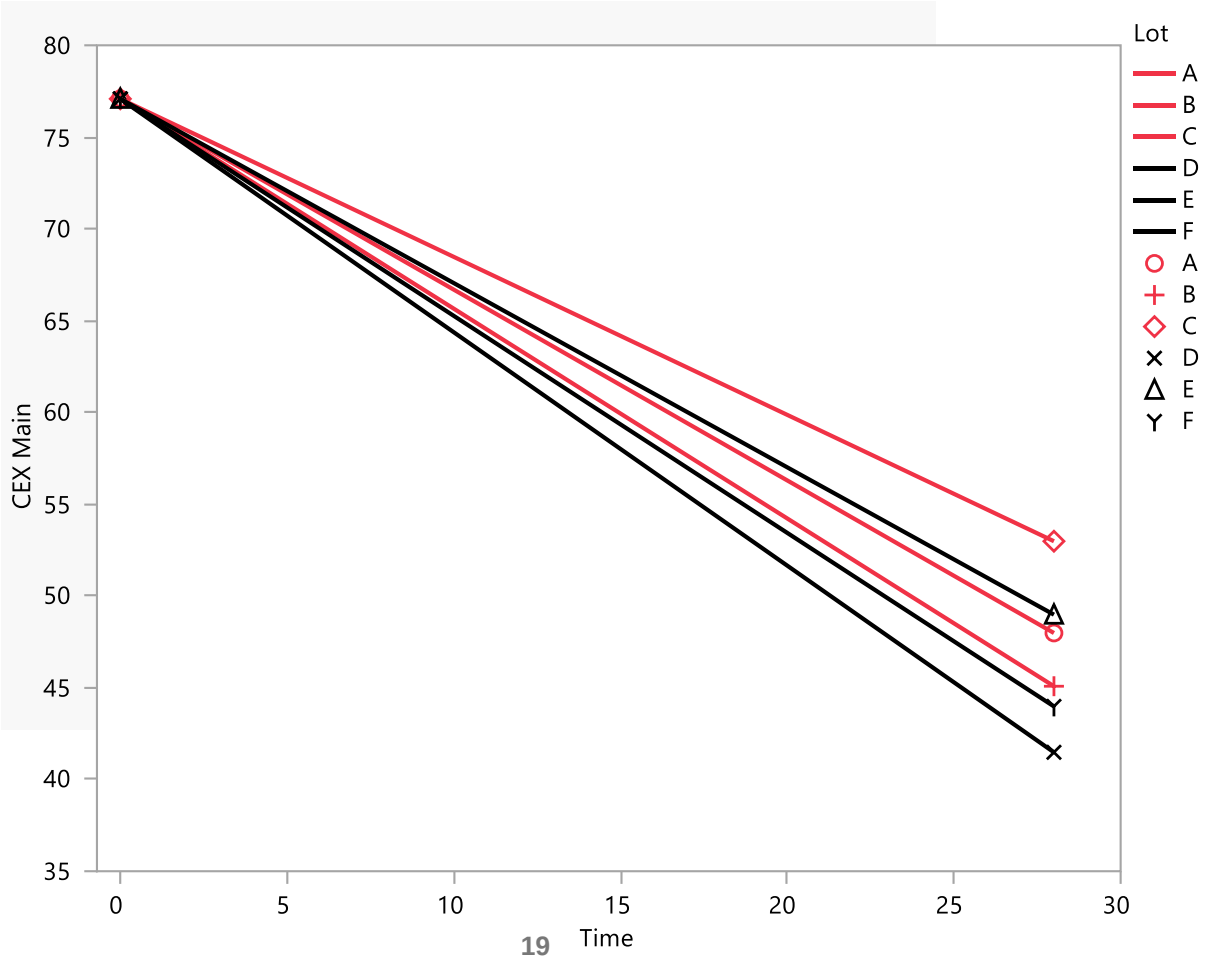


Calibration of Rule 3 (Quality range)

- Note that *K* for Rule 3 can be calibrated to always give the desired type 1 error rate for any given value of ES that defines comparability.
- For example, with 3 lots, if you desire a type 1 error rate of 0.10 at ES=3, the required value of *K* is 2.33.
  - Type 1 error rate with *K*=3 was **0.231**
  - This gives a power of **0.709** at ES=0 compared to Rule 2 power of **0.900** with type 1 error rate of 0.11. (Rule 2 best choice with 3 lots)
- Simulations to compute these numbers are easy to perform---see your friendly statistician.

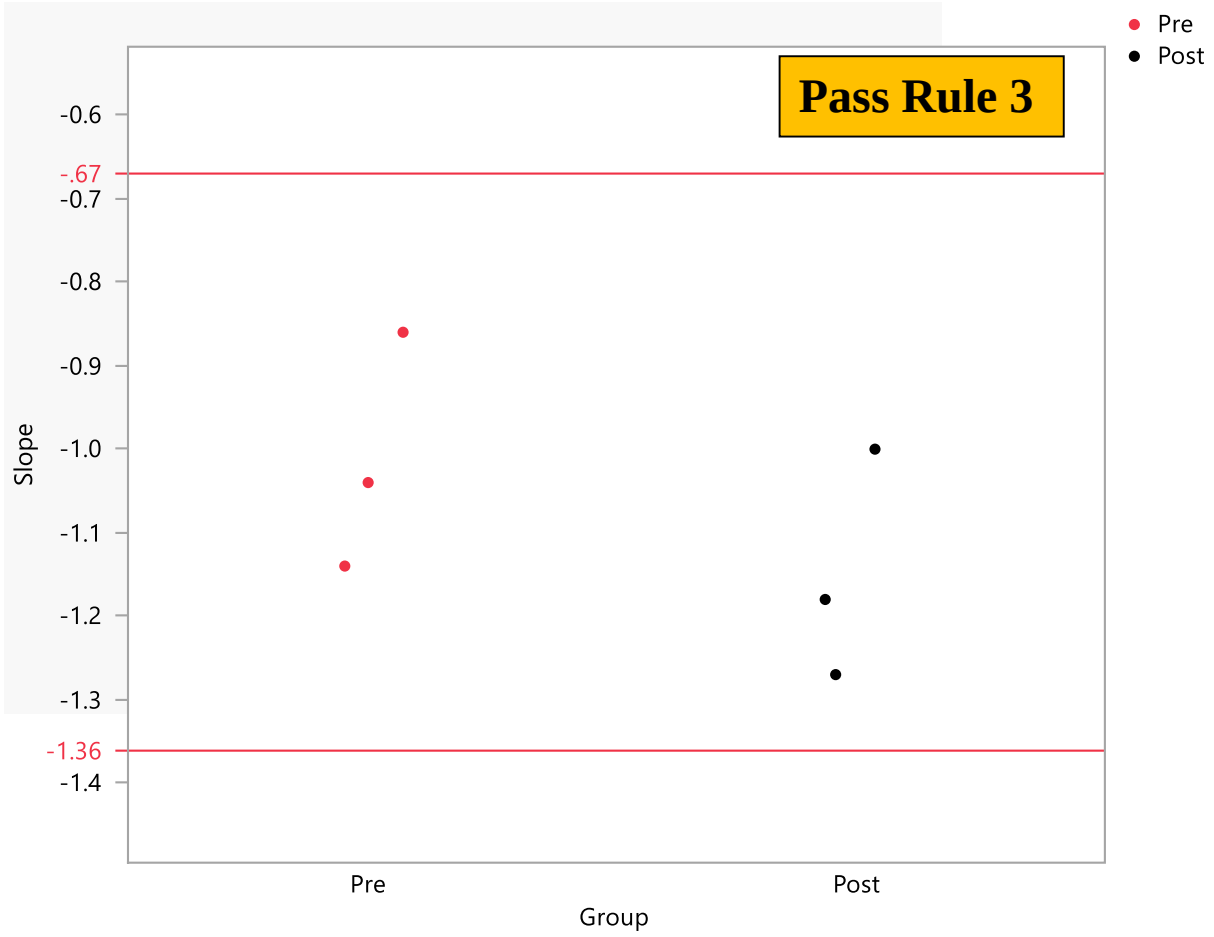
Our example analyzed-Rule 1 and 2

- We fail Rule 1 (all in range) and pass Rule 2 (at least one in range)



Our example analyzed-Rule 3 and ES

- Estimated ES=1.2
- Quality Range:



Summary

- Visual interpretation or computation of a quality range can be used to assess differences in stability slopes.
- One needs a definition of comparability (in terms of ES) to allow proper selection of a decision rule.
- Definitions of comparability need to consider criticality of the attribute.
- Take-home rule of thumb: *If you can accept ES=3 as definition of comparability*, you can claim comparability if at least one of three post-change lots overlaps range of three pre-change lots (Rule 2).

Backups

Probabilities of passing with three rules

Type 1 error rate if ES=3 defines comparability

Number of lots	Effect size	Rule 1: Probability all in min max range	Rule 2: Probability at least one in min max range	Rule 3: Probability all in quality range K=3
3	0	0.200	0.900	0.824
6	0	0.228	0.998	0.861
3	1	0.120	0.738	0.712
6	1	0.110	0.962	0.700
3	2	0.027	0.377	0.464
6	2	0.015	0.717	0.364
3	3	0.002	0.111	0.231
6	3	0.000	0.295	0.114

We want a high probability of passing for ES=0, and low probability for ES=3

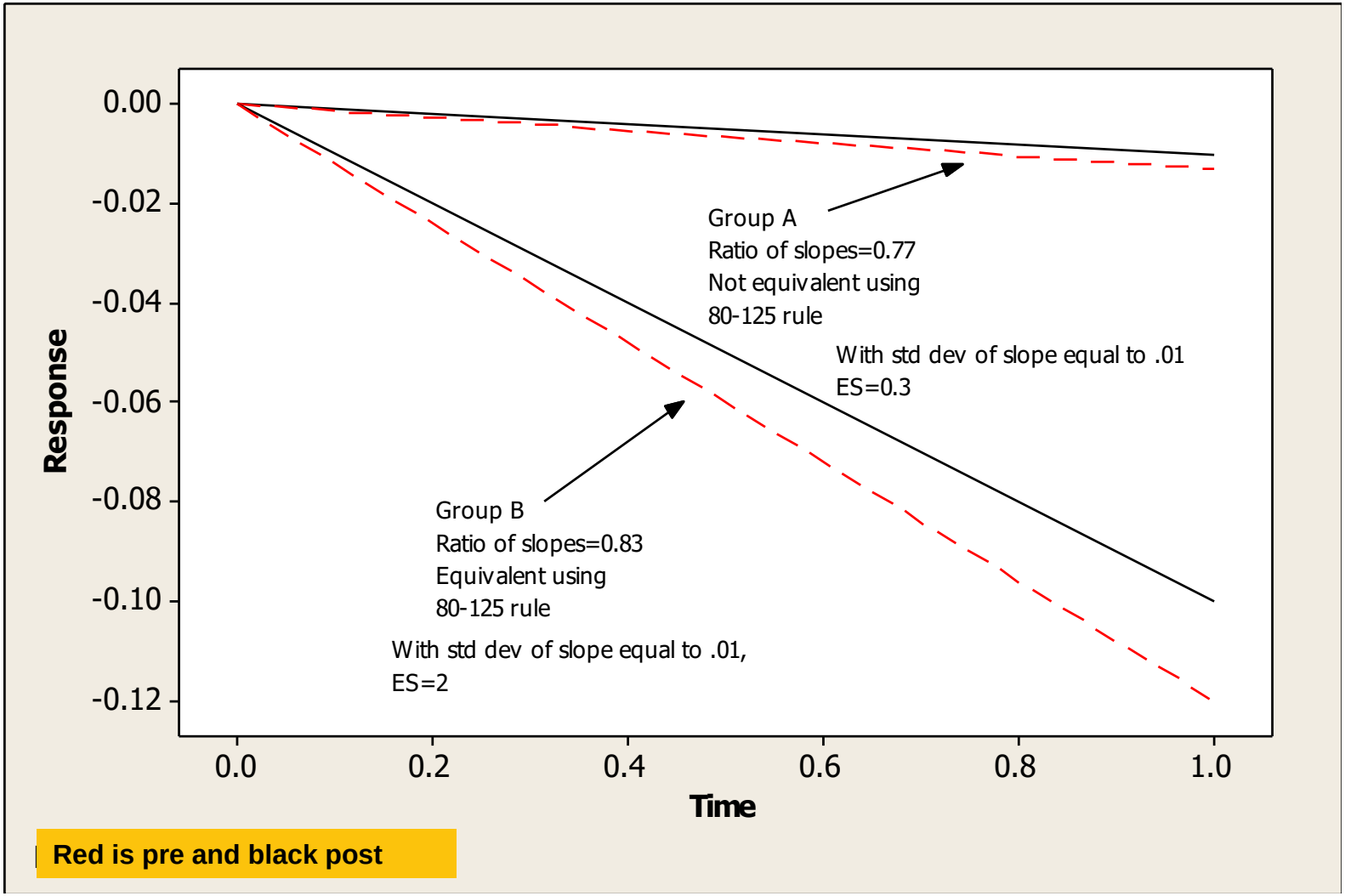
Rule 2 seems to be best rule with 3 lots—low type 1 error rate (0.109) and high probability of passing with ES=0 (0.898). However, if you require ES<3 as a definition of comparability, it will not be acceptable.

Rule 3 allows for calibration of K so one can obtain desired type 1 error rate for any ES used to define comparability.

Why use difference of slopes instead of ratio of slopes?

- Difference in slopes has a meaningful unit of measure for SME.
  - Describes change difference in the quality attribute over a fixed period of time.
  - A ratio of slopes has no unit of measure or meaningful interpretation.
- Often cited 80%-125% rule for ratio of slopes originally used in bioequivalence studies has no meaningful interpretation for the present problem.
  - In bioequivalence study, this range was selected to ensure that at least 80% and no more than 125% of the ingredient is absorbed in the same time period.
- Ratio of slopes is not always consistent with visual representation and cannot be defined if slopes close to zero have different signs.

Ratio of slopes is not always consistent with visualization



Statistical tests of equivalence

- Burdick and Sidor (2013) have proposed a statistical test of equivalence of slopes.
- However, equivalence tests ignore differences in variability that can be considered with the heuristic tests mentioned today.

References

- Burdick, R. K., and Sidor, L. (2013). “Establishment of an equivalence acceptance criterion for accelerated stability studies”, *Journal of Biopharmaceutical Statistics*, 23:4, 730-743.
- Lei, L., and Olson, K. (2010) “Evaluating Statistical Methods to Establish Clinical Similarity of Two Biologics”, *Journal of Biopharmaceutical Statistics*, 20: 1, 62 — 74.