Bioequivalence Concepts and Data Analysis

Marilyn N. Martinez, Ph.D.

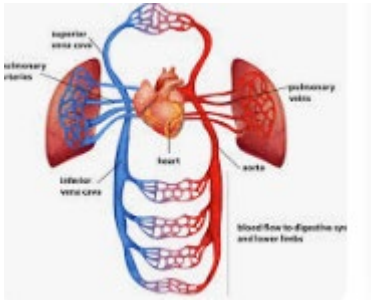
Chair, VICH Bioequivalence Expert Working Group

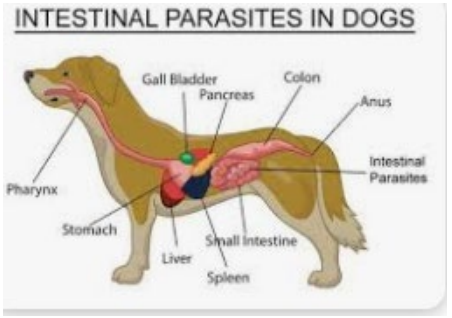
Information to be Covered

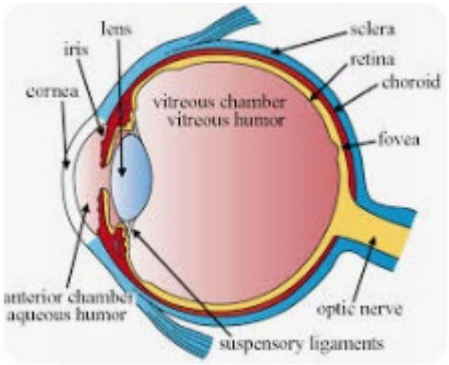
- What is bioequivalence (BE)?
- How to establish treatment BE based on blood level profiles.
- BE study design and analysis: why we need statistics?
  - Selecting the reference product and lot
  - Prandial state: why fasted?
  - Why do we still need residual depletion information if two products are BE based on blood level data?
  - Crossover versus parallel design.
  - Why is dose normalization generally not permitted
  - What dose should be used?
  - Importance of having an adequate washout period.
  - Need for a BE study in each major species on reference label.
- Mathe-magics
- Basic statistics used in the determination of product BE
- Example of how to analyze a 2-treatment, 2-period, 2-sequence crossover
- Summary of key points.

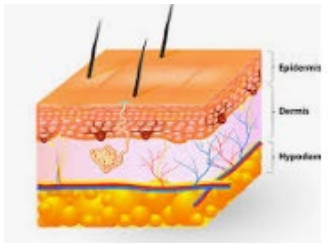
What is Bioavailability?

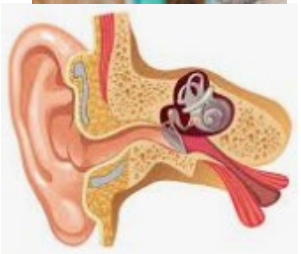
Bioavailability is a measure of the rate and fraction of the dose that successfully reaches either the drug's intended site of action or the bodily fluid domain from which the drug's intended targets have access.











Drug Bioavailability, Price and Patel, 7/20/2023.
https://www.ncbi.nlm.nih.gov/books/NBK557852/

What is Bioequivalence (BE)?

The absence of a difference (within predefined acceptance criteria) in the bioavailability of the active pharmaceutical ingredient (API) or its metabolite(s) at the site of action when administered at the same molar dose under similar conditions in an appropriately designed study.

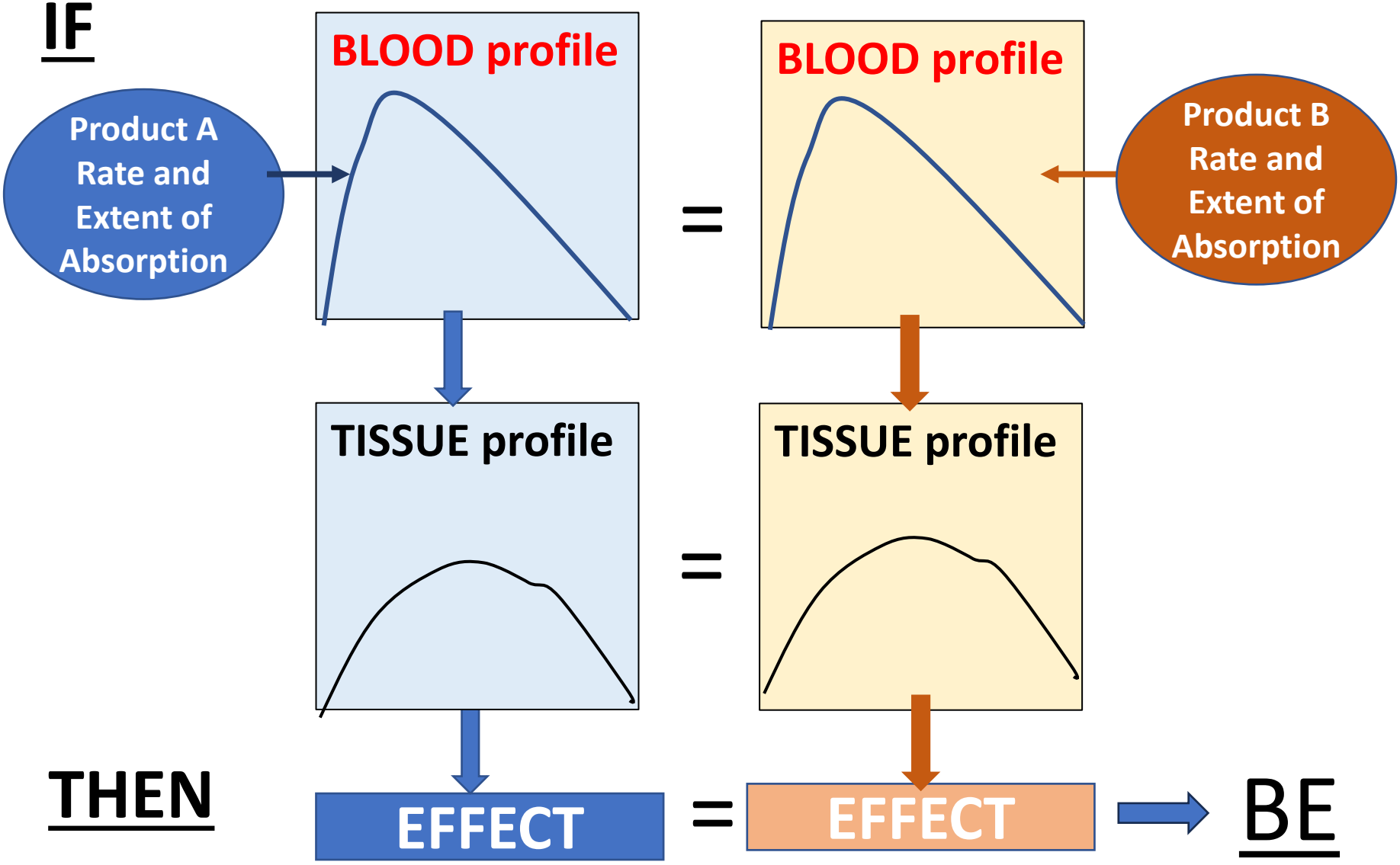
The remainder of this presentation will focus on the assessment of BE for systemically absorbed drug products (see VICH GL52).

What is Bioequivalence (BE)?

Plasma concentrations provide a surrogate for demonstrating product equivalence in terms of safety and effectiveness.

The assumption is that if two products are bioequivalent, their in vivo performance (safety and effectiveness) will be indistinguishable to the patient and are therefore freely interchangeable.

Theoretical Basis Behind BE Concepts



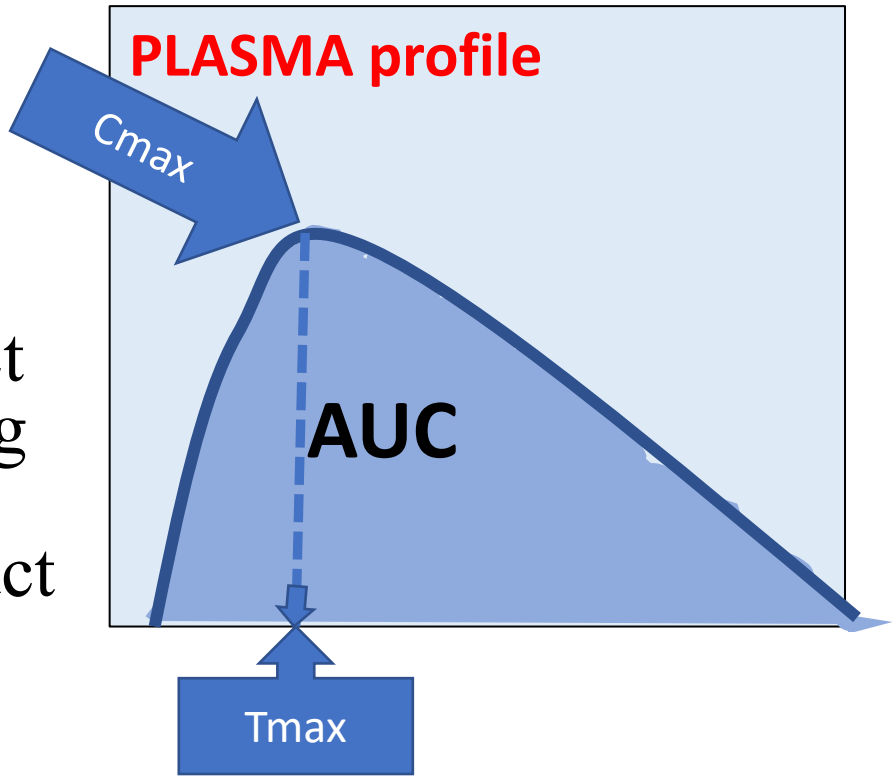
What is Bioequivalence (BE)?

- Errors can occur when using pharmacokinetic (PK) models to fit the observed data and then apply these models to describe the rate and extent of drug absorption. Therefore, we typically rely on the use of a noncompartmental analysis (NCA) to assess product comparability.
- Pivotal NCA parameters include the observed peak exposure (Cmax) and total drug exposure (area under the concentration vs time curve, AUC) from time zero (predose) to the last sampling time when drug concentrations are $\geq$ the analytical limit of quantification (LOQ). This is often referred to as AUC0-last.
- Samples that drop below the LOQ are not included in the calculation of AUC0-last.

How to Establish Equivalence of Blood Level Profiles

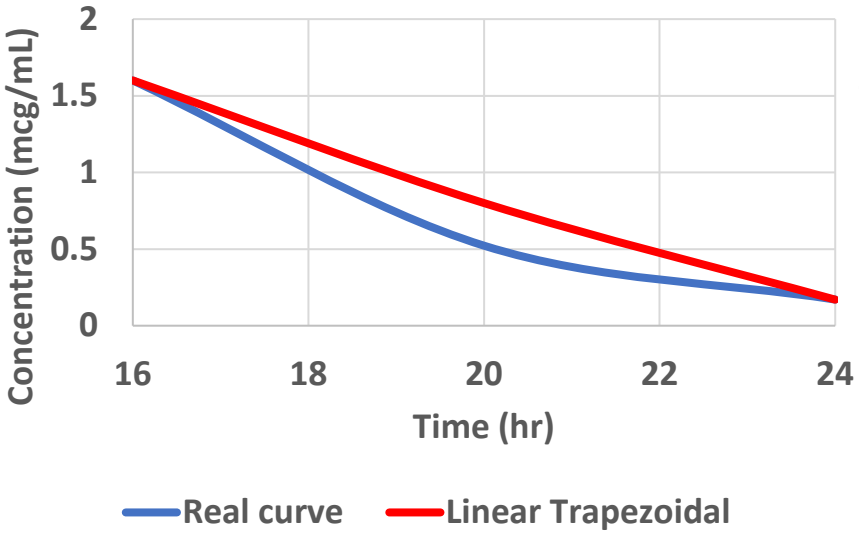
- Observed peak exposure (Cmax)
- Time to peak (Tmax)
- Total exposure (AUC)

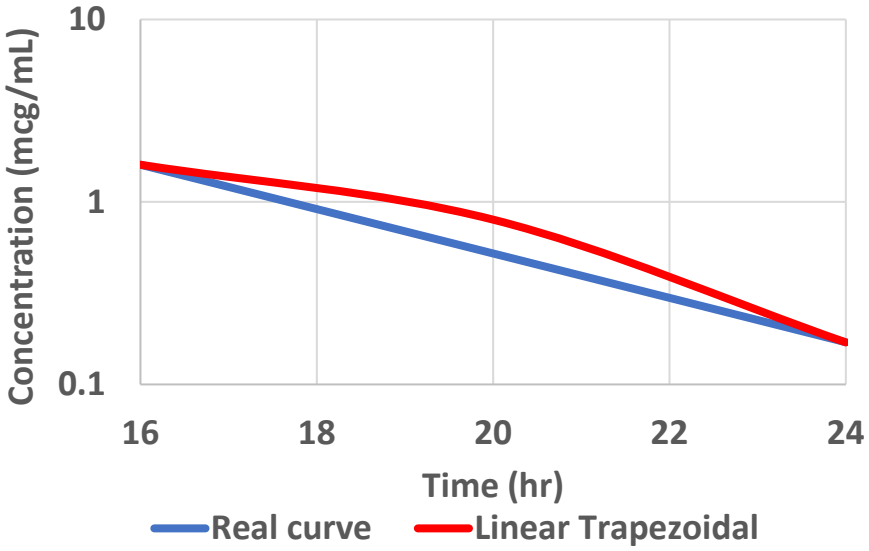
BE study blood sampling schedule is based upon the PK of the reference product with the goal of minimizing the difference between the true versus observed product values of AUC, Cmax and Tmax.



Which is Better: Linear or Linear-Ln AUC Estimate?

Since the depletion portion of the curve typically exhibits an exponential decline, the use of the linear trapezoidal rule during the depletion phase can introduce error in our AUC estimation.





Which is Better : Linear or Linear-Ln Estimate of AUC?

- **Linear trapezoidal rule**: Summation of trapezoids

$$\frac{1}{2}(C_1 + C_2)(t_2 - t_1)$$

- **Linear-Ln trapezoidal rule**: Summation of Linear trapezoidal rule from time zero to Tmax and sum of Ln trapezoids determined after Tmax. The trapezoids after Tmax are calculated as:

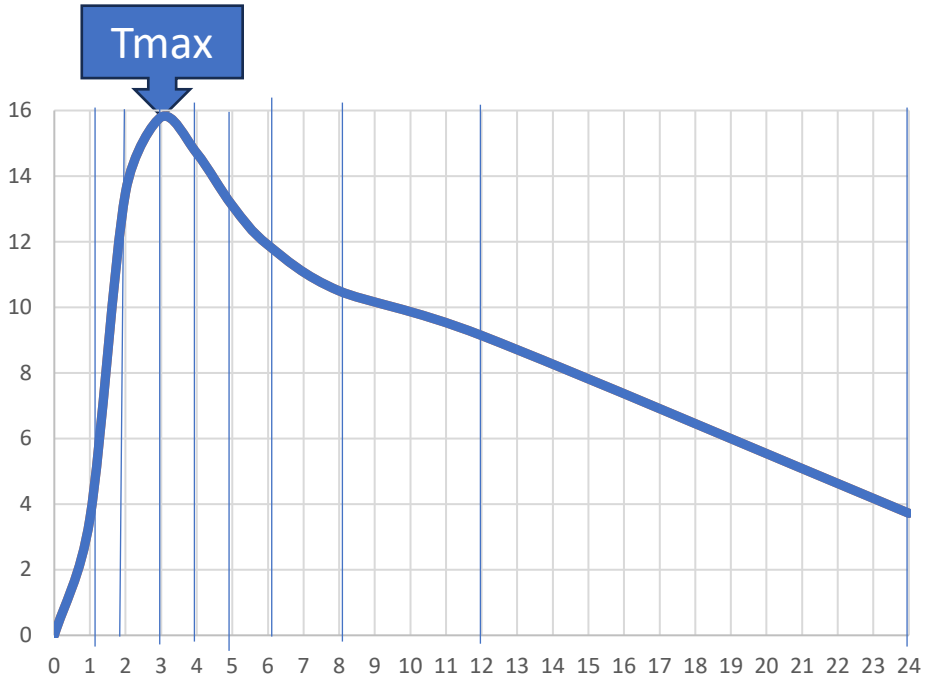
$$\frac{C_1 - C_2}{ln(C_1) - ln(C_2)}(t_2 - t_1)$$

where t_2 - t_1 is the duration of a time interval and C_2 and C_1 are the concentrations at t_2 and t_1.

https://www.certara.com/knowledge-base/calculating-auc-linear-and-log-linear

Which is Better: Linear or Linear-Ln Estimate of AUC?

Example:

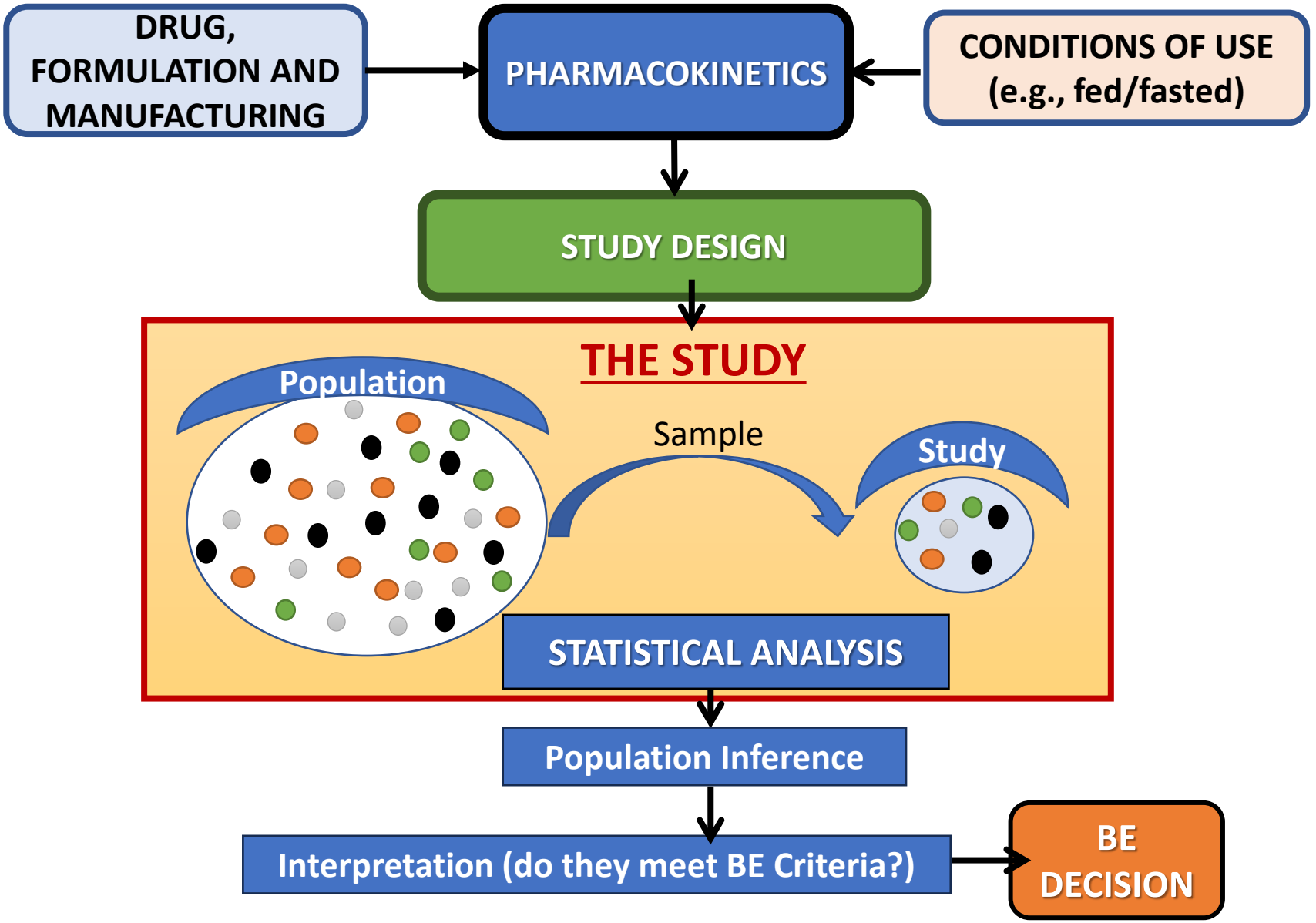


Time	Linear Trap Rule	Linear-Log Trap
0		
1	1.78	1.78
2	10.31	10.31
3	24.96	24.96
4	40.21	40.20
5	54.11	54.09
6	66.61	66.58
8	89.01	88.95
12	128.29	128.17
24	205.45	200.52

Differences are primarily at end of the sampling scheduled when the sampling times are furthest apart.

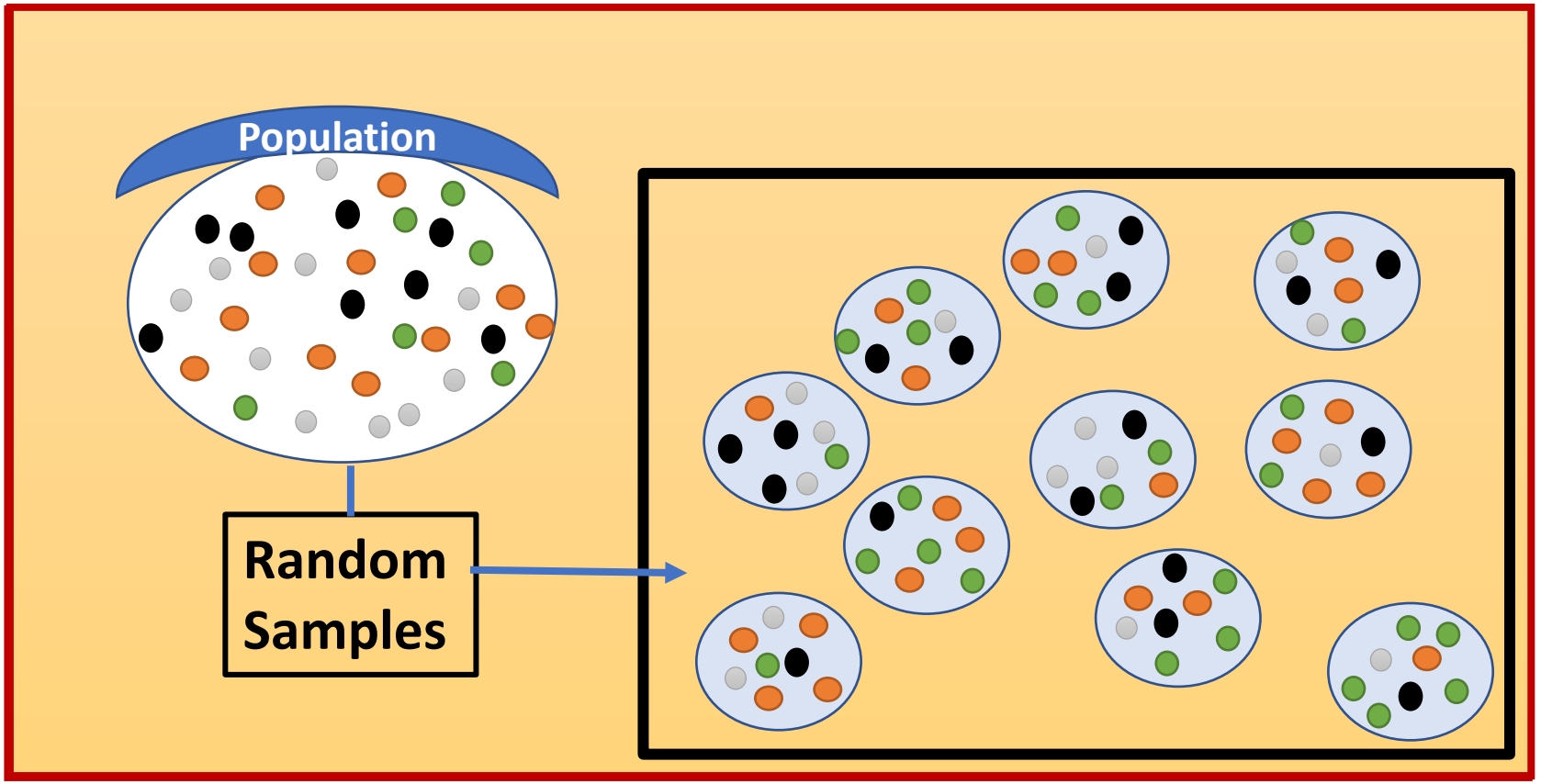
See Exercises Case Study 1

BE Study Design and Analysis: an integration of disciplines.



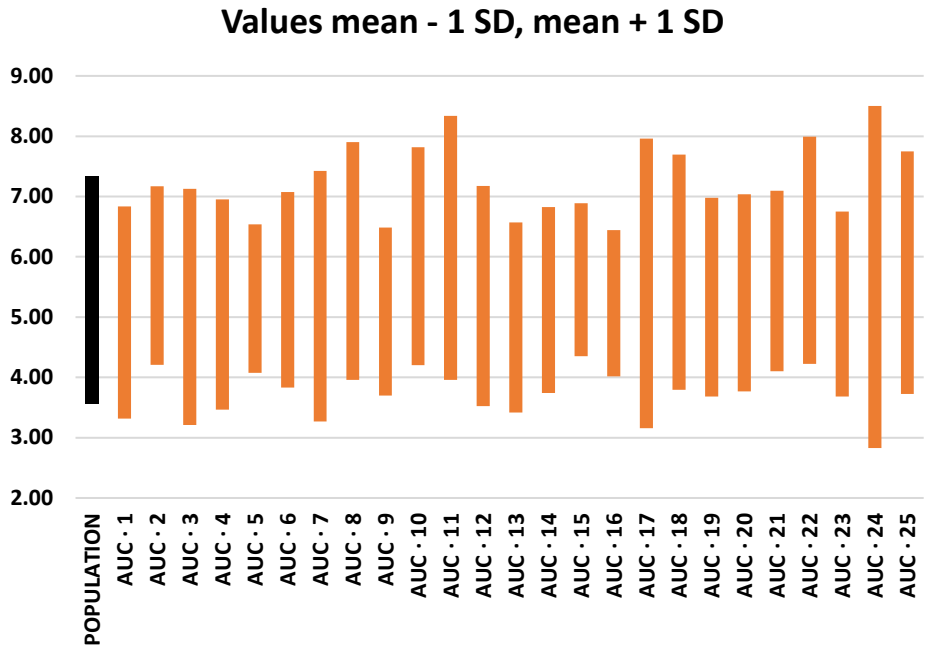
BE Study Design and Analysis: Why we need statistics

If we repeatedly collect random samples from the same population, we will obtain a distribution of outcomes. Statistics provides a mechanism for generating inferences about the actual patient population from a sample obtained from the population.



BE Study Design and Analysis: Why we need statistics

For example, Monte Carlo simulation was used to generate 25 sets of samples containing 25 observations. Compared to the "population" (1200 simulated AUC values), we can see the distribution of values (mean – 1 SD, mean +1 SD) can be higher or lower than that of the total population. Without statistical methods for generating population inferences, we could generate incorrect conclusions about the actual distribution AUC in that population.



The 90% Confidence Interval

You will see this slide again later

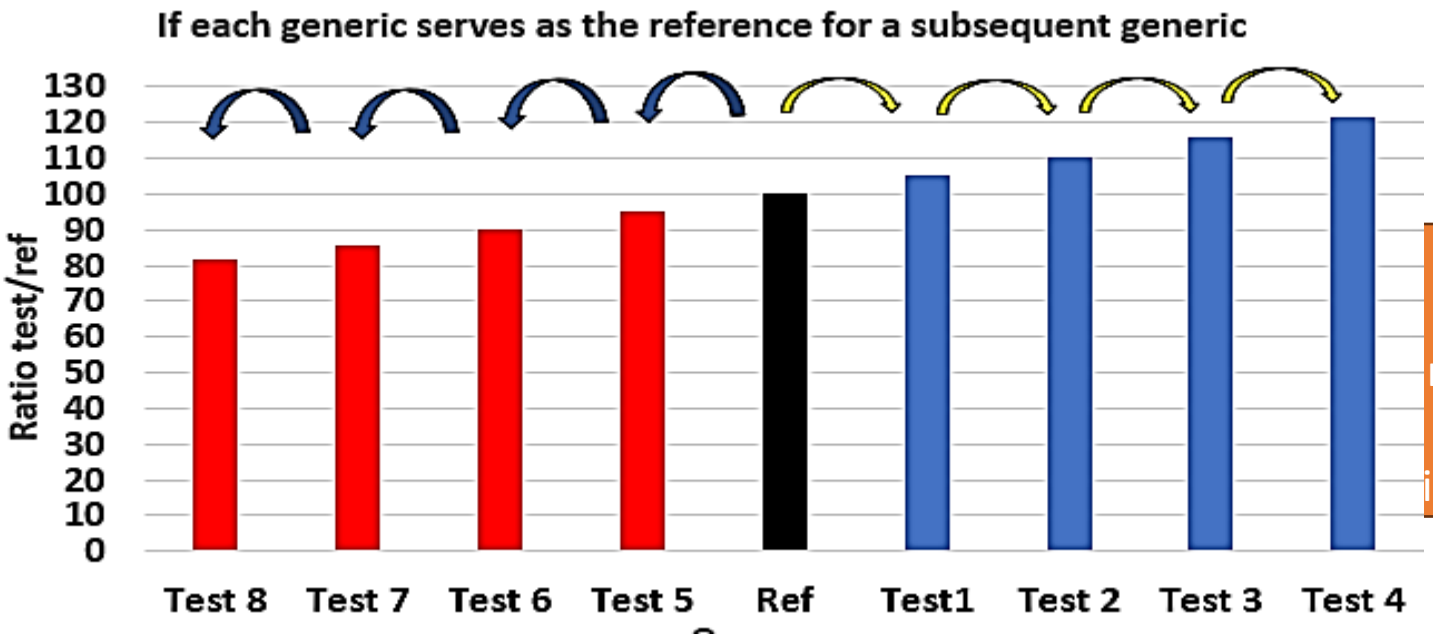
90% Confidence interval for an estimated parameter *(e.g., mean AUC)*: if the same population is sampled on numerous occasions and interval estimates are made on each occasion, the resulting intervals would bracket the true population parameter (i.e., the T/R ratio) in approximately 90% of the cases.

the *NIST/SEMATECH e-Handbook of Statistical Methods* (April, 2012), NIST https://www.itl.nist.gov/div898/handbook/prc/section1/prc14.htm

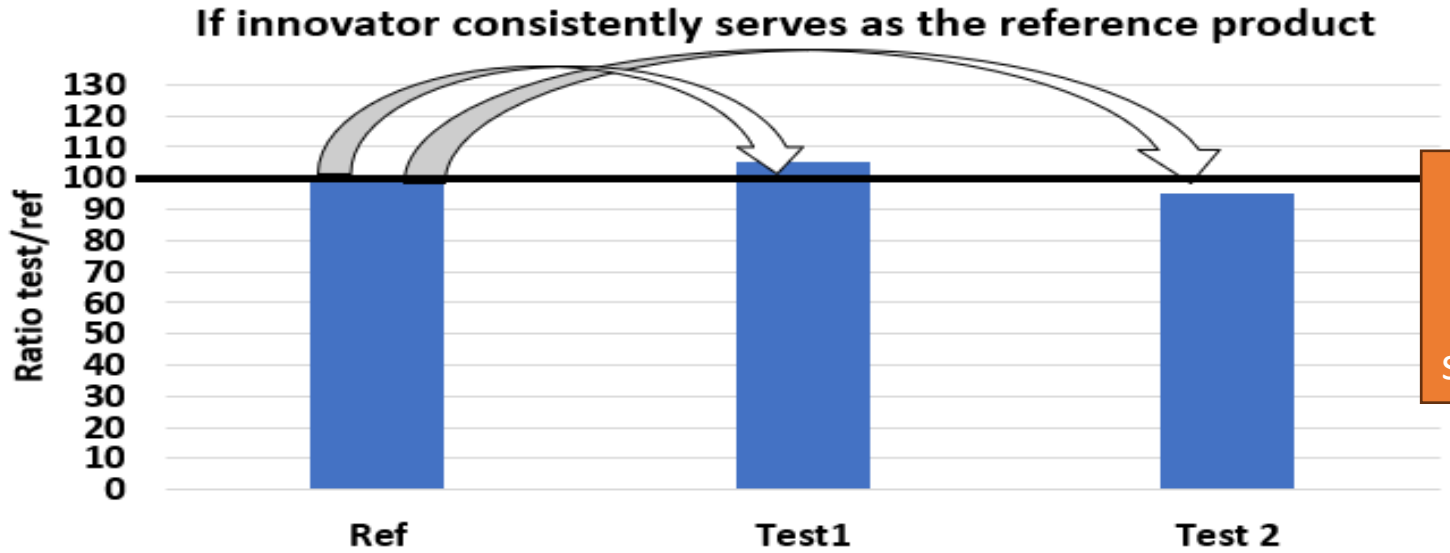
Why Do We Need to Always Use the Same Reference Product?

By constraining all BE studies to the same reference product, we avoid the between-generic drift that can occur if different products are selected as the reference (e.g., if some studies used an approved generic as reference instead of the marketed innovator).

Why Do We Need to Always Use the Same Reference Product?



The generic differs from its reference by 5% - risk loss of interchangeability



Each generic differs by 5% relative to the same Ref product.

VICH GL52: Criteria for Selecting Products to be Tested

The API content of the test and reference products should be assayed prior to conducting the BE study. **To be internationally acceptable**, it is recommended that the assay content of the **batches from which test and reference products were obtained should differ by no more than ±5% from each other**.

This caveat reduces the possibility of a product batch with higher or lower potency influencing conclusions about product BE.

Simulated Example

- Each tablet is supposed to contain 10 mg.
- Each dog is administered one tablet.
- Batch 1 of the test product had the same potency as the reference. Batch 2 had a 30% overage.
- The bioavailability of the test product was 25% less than that of the reference.

	Ref	Test batch 1	Test batch 2
DOSE (mg)	10	10	13
F	1	0.75	0.75
CL (mL/hr)	20	20	20
AUC (µg*hr/mL)	500	375	487.5
T/R		0.75	0.975

Camouflages product inequivalence

BE Study Design: Prandial State

For orally administered canine and feline drug products, studies should be conducted in fasted animals unless the approval for the reference product recommends administration in the fed state only, in which case the study should be conducted accordingly.

Why Conduct Studies in the Fasted State?

The fasted state tends to be more sensitive to differences in product rate and extent of absorption.

For example, food-induced slowing of gastrointestinal (GI) transit allows more time for products to dissolve. This slowing can camouflage differences in product in vivo dissolution rates that might occur in fasted animals.

Food-induced release of bile salts and changes in gastric pH can affect drug solubility, facilitate drug dissolution, and minimize formulation differences that might occur in the fasted state.

Example of BE Fed, Not BE Fasted

Human BE Study involved two different oral tablet formulations under fed and fasted state

	Bioequivalence of test vs. reference product under fasted conditions					
	Intra-subject CV (%)	Geometric LSMeans		Ratio (%)	90% Confidence limits (%)	
		T-fasted (n=33)	R-fasted (n=33)		Lower	Upper
C_{max} (ng/mL)	15.1	445.68	275.99	161.49	151.61	172
AUC_{0-T} (ng·h/mL)	11.5	11,444.3	8246.08	138.78	132.28	145.61
$AUC_{0-\infty}$ (ng·h/mL)	11.7	11,684.54	8487.3	137.67	131.11	144.56
	Bioequivalence of test vs. reference product under fed conditions					
	Intra-subject CV (%)	Geometric LSMeans		Ratio (%)	90% Confidence limits (%)	
		T-fed (n=33)	R-fed (n=33)		Lower	Upper
C_{max} (ng/mL)	13.6	380.61	390.74	97.41	91.95	103.19
AUC_{0-T} (ng·h/mL)	7.5	11,279.27	11,574.23	97.45	94.41	100.6
$AUC_{0-\infty}$ (ng·h/mL)	7.7	11,544.87	11,892.18	97.08	93.97	100.29

Dayan et al., Infect Dis Clin Microbiol 2023; 5(4): 341-52

But Why Does Human Medicine Sometimes Ask for Both Fed and Fasted BE Studies?

- Concerns include the potential for drug product excipients to interact with food or to influence how the product behaves in the fed state.
- However, it is relatively rare for a fasted study to meet BE criteria but fail in the fed state (estimated to be ~3% for immediate release formulations). (*AAM, Your Generics & Biosimilar Industry, May 01, 2019*).

Why Plasma Concentrations Cannot Insure the Same Withdrawal Period?

A Simulated Example of a Subcutaneous Injection:

In this SIMULATED example, **97%** of the test product is absorbed at an identical rate as the reference.

However, **3%** of the injection is absorbed at a much slower rate.

100% of both products are ultimately absorbed.

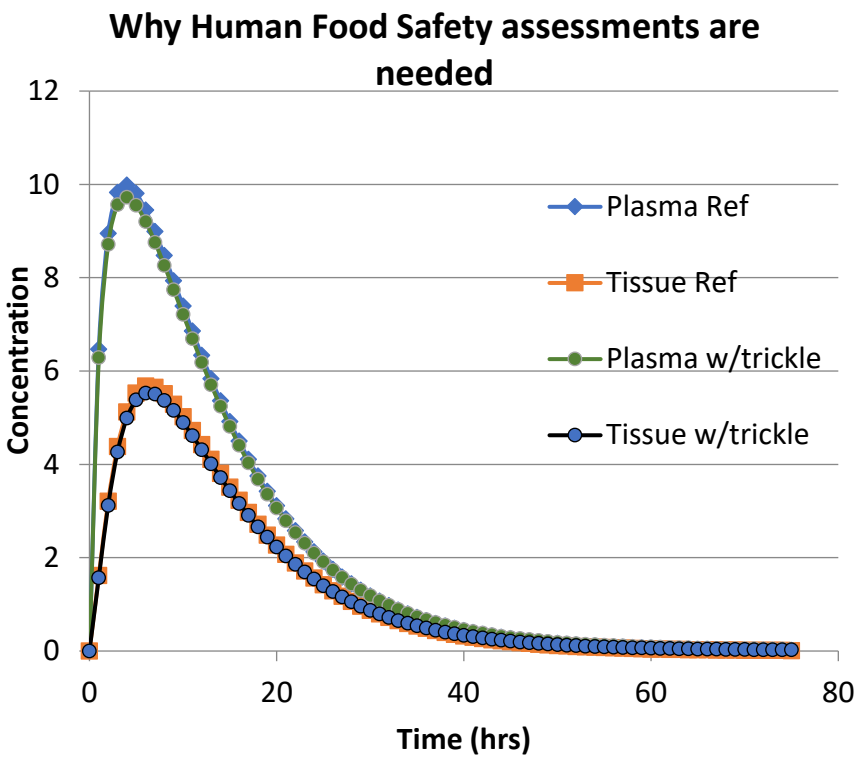
Results:

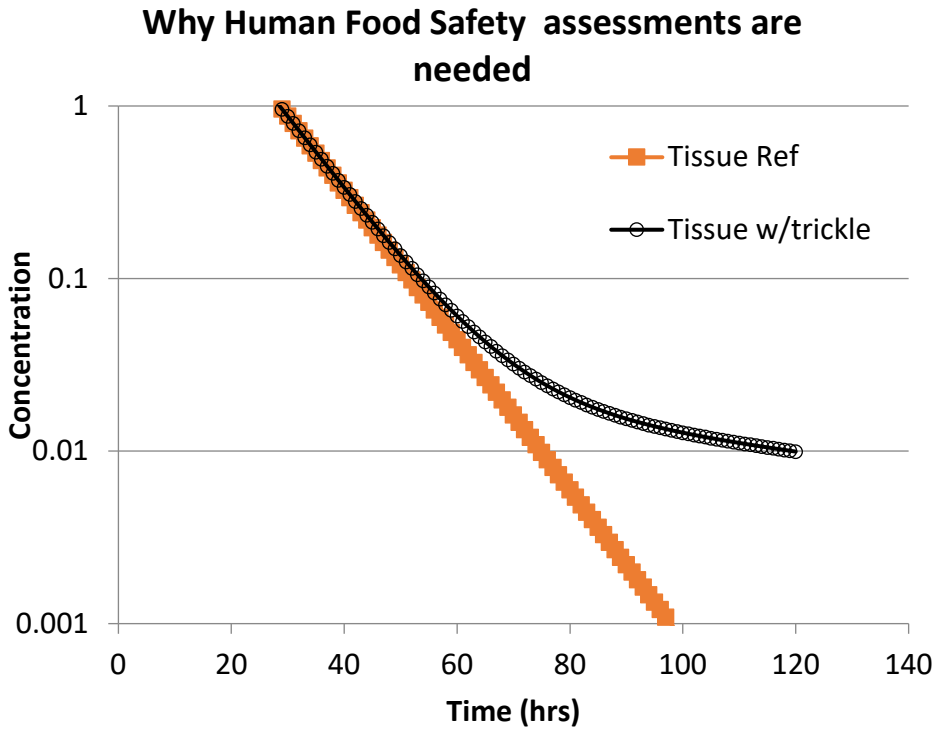
Plasma AUC_{48} Test/Ref ratio = 0.98 → highly likely to pass BE

Plasma Cmax Test/Ref ratio = 0.974 →highly likely to pass BE

Plasma Tmax = 4 hrs for both test and reference products

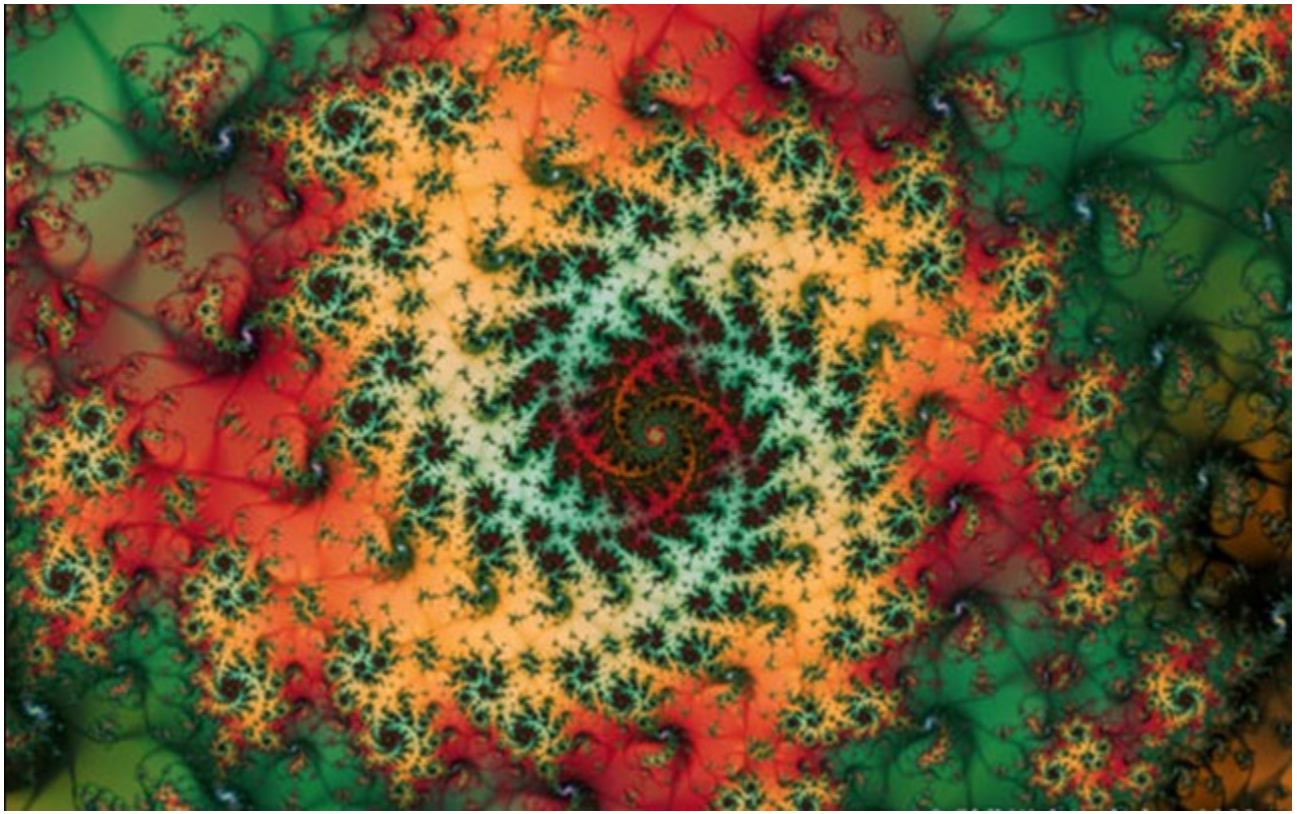
Tissue Time to MRL (0.01 units) = **75** hrs (ref) and **120** hrs (test)





The small proportion of drug released by trickle absorption from the test product did not impact the determination of blood level BE but it did influence the time to the established tolerance/MRL (i.e., the withdrawal period).

Study Design



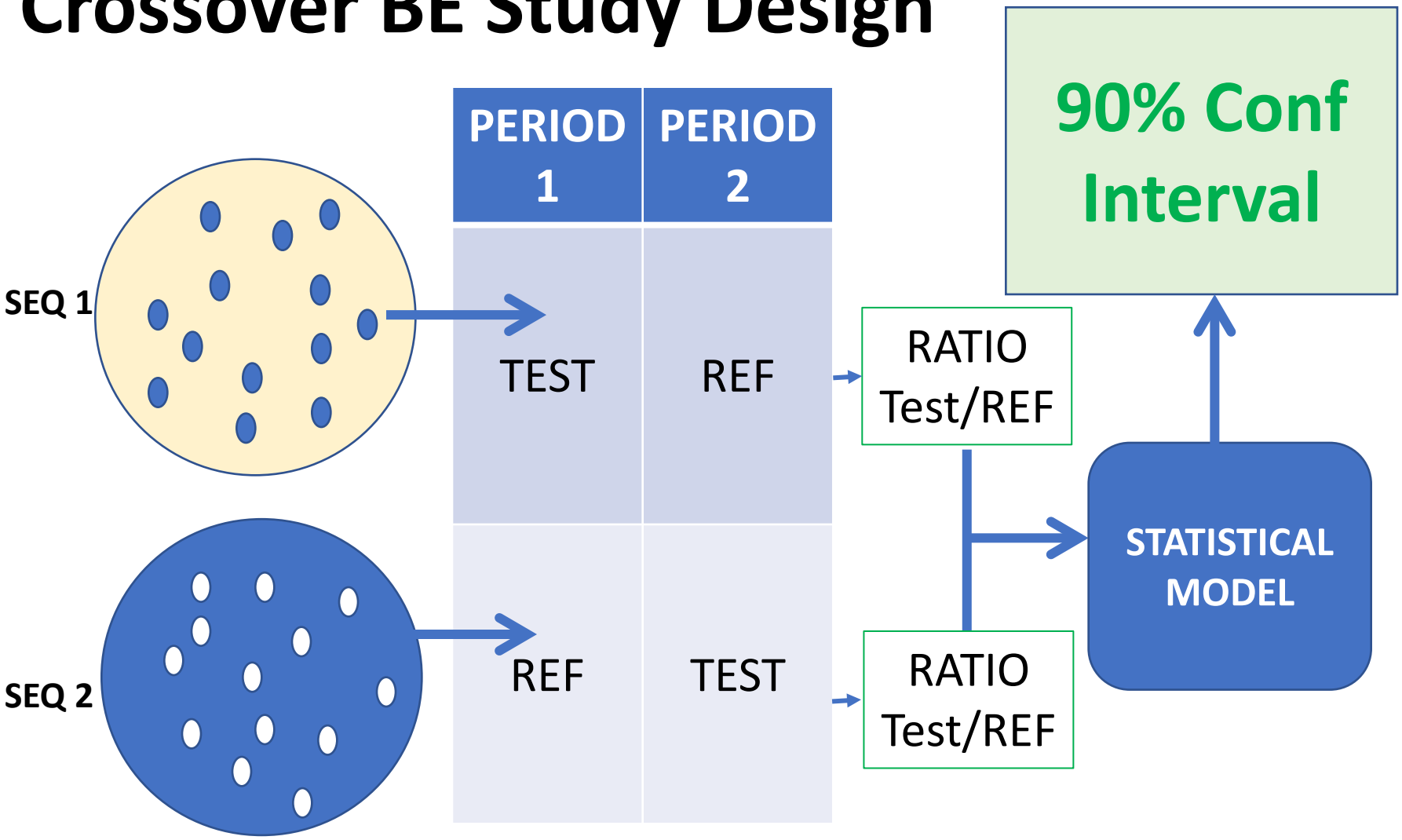
When feasible, it is best to conduct the BE study using a crossover study design.

Parallel study: width of confidence interval is based on "within + between" subject variability.

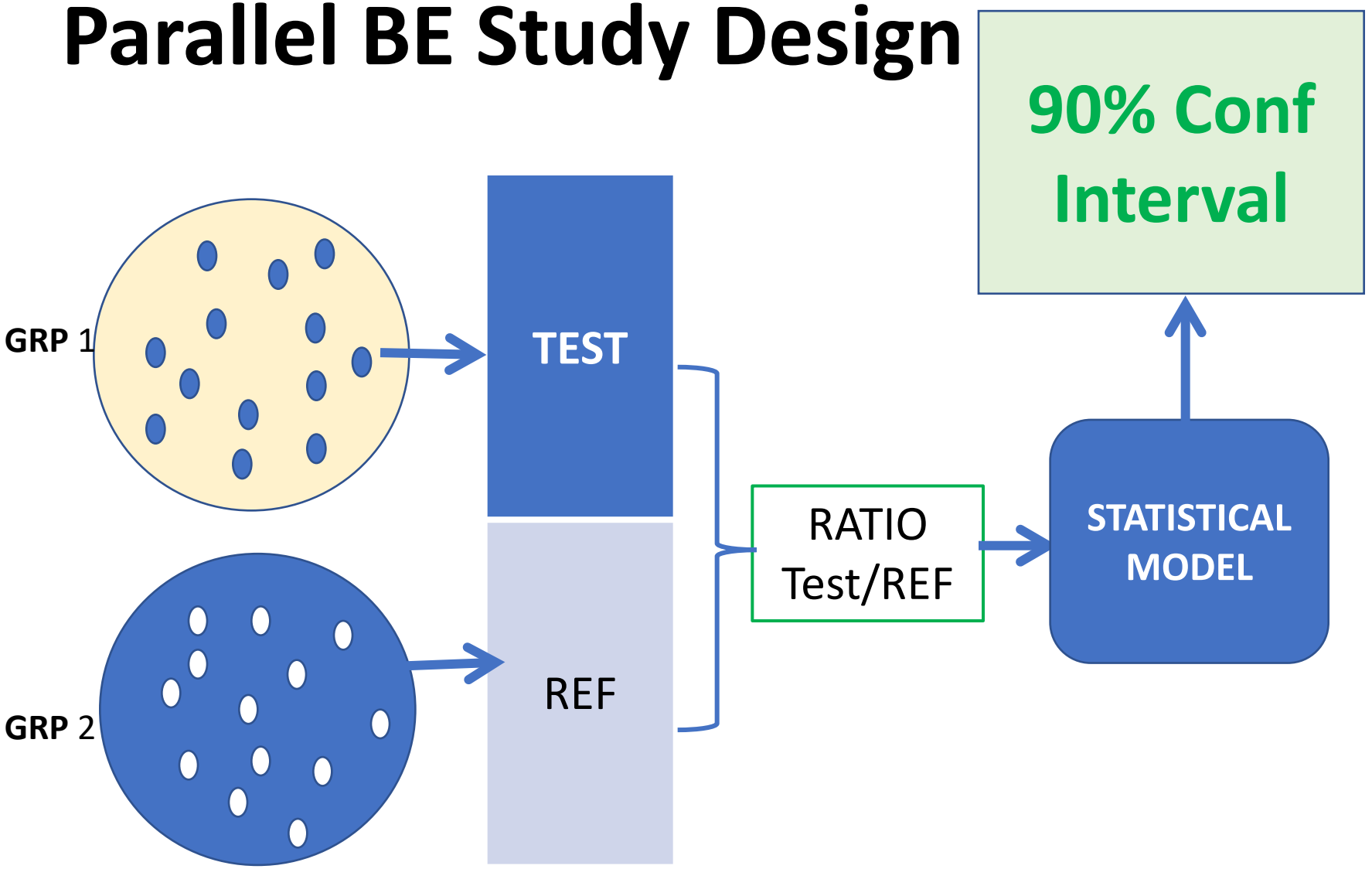
Crossover study: each subject is compared to itself. Therefore, the variability associated with crossover studies are typically less than that of parallel studies.

Both study designs can contain other sources of variability not explained by the statistical model (such as analytical and experimental error).

2-Sequence, 2-Period, 2-Treatment Crossover BE Study Design



Parallel BE Study Design



When is a Parallel Study the Preferred Design?

- When there is the risk of a physiological carryover (e.g., enzyme induction) that could influence period 2 systemic drug concentrations.
- When the time needed to ensure the absence of a carryover effect is not feasible due to:
  - Formulation-independent effects: the PK of the drug itself results in a very slow elimination that is not related to formulation (e.g., cefovecin).
  - Formulation-dependent effects: the rate of drug release is intentionally modified (e.g., extended-release products) and drug release can last from months to years.
- When the necessary duration of a washout risks significant physiological changes (e.g., animal growth).
- When the target species has insufficient blood volume to allow for a two-period study design (e.g., fish).

- *See Exercises, Case Study 2*

Why Dose Normalization is Not Needed

Unlike most human medications, veterinary medications are typically administered on a mg/kg basis. Therefore, small differences in body weight (BW) are not critical.

But what would happen if only one tablet strength were available, and the subjects reflected a range of body weights? ***And what if the study were conducted using a crossover versus parallel study design?***

Why Dose Normalization is Not Needed

In this simulated example, we consider two subjects using a parallel vs crossover study design.

A 10 mg dose is administered to all subjects.

Animals weigh either 7 kg or 10 kg. In all cases, the test and reference products are associated with the same fraction of dose absorbed.

Does the difference in BW influence the Test/Reference AUC ratio?

Why Dose Normalization is Not Needed for a Crossover Study

	Subj	Product	Body Wt (kg)	Dose	AUC	T/R
Crossover	1	Test	7	10 mg	816	1
	1	Ref	7	10 mg	816	
	2	Test	10	10 mg	574	1
	2	Ref	10	10 mg	574	
Parallel	1	Test	10	10 mg	574	0.7
	2	Ref	7	10 mg	816	

For a crossover study, comparisons are generated within subjects. Therefore, differences in BW do not impact the T/R ratio.

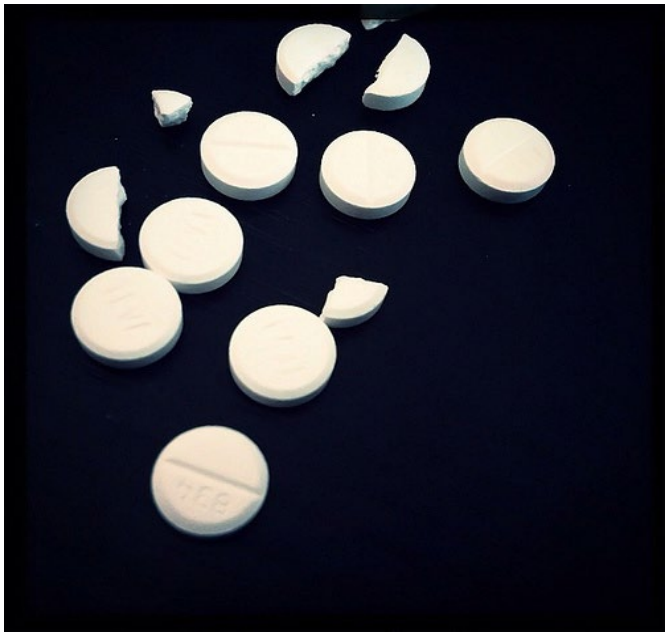
For a parallel study, differences in BW could be a problem if the same mg dose is administered regardless of body weight.

Why Dose Normalization is Not Needed

When a parallel design is used and the test and reference products cannot be administered on a mg/kg basis, an investigator can minimize between-subject variability by selecting subjects of similar body weights.

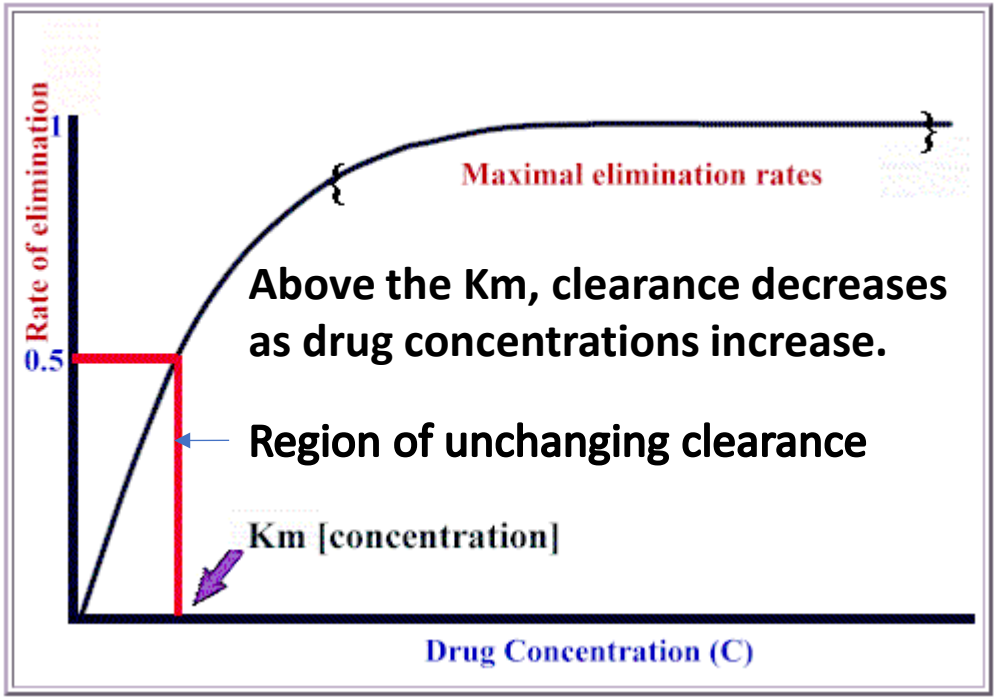
Point of Note

Shaving tablets is **NEVER** considered appropriate because this alters the product geometry, its outer coating, and can alter drug release.



Which Dose Should We Use?

We typically recommend using the highest approved dose because of potential changes in product relative exposure if there are saturable elimination processes (Michaelis-Menten kinetics, where maximum elimination, Vmax, is in moles/unit time).



With saturable kinetics, clearance (L/hr) is not constant but rather varies as a function of concentration!

https://www.pharmacology2000.com/General/Pharmacokinetics/kinobj1E.htm

If rate of elimination flattens as dose increases, products may be BE at the lower dose but inequivalent at the higher dose (different concentrations leading to differences in the degree of saturation).

Example (simulated in Simcyp)

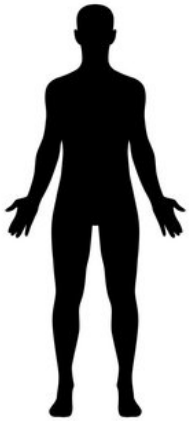
In all cases, the bioavailability (F) of the test product was simulated to be 20% lower than that of the reference.

At doses within the range associated with linear kinetics, the AUC T/R ratio was 0.80 (keep in mind that AUC=(Dose×F)/CL.

However, when the doses were in a range associated with saturable elimination processes, the T/R AUC ratios dropped to 0.75. This is because the lower concentration of the test product had higher CL than the reference.

Doses tested, both with same F	AUC T/R
Below Saturation	0.8
Within Region of Saturation	0.75

In human medicine, dosing at highest label dose avoids risk of inequivalence at the high dose that may be missed at lower doses.



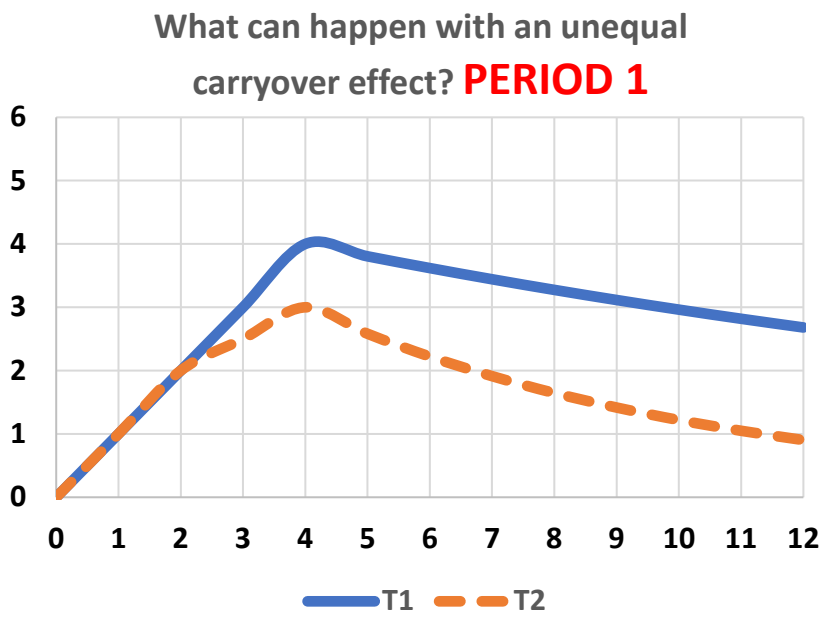
In veterinary medicine, since we typically dose on a mg/kg basis, the issue of the highest label dose is relevant only if the product is approved for use at more than one specific mg/kg dose.

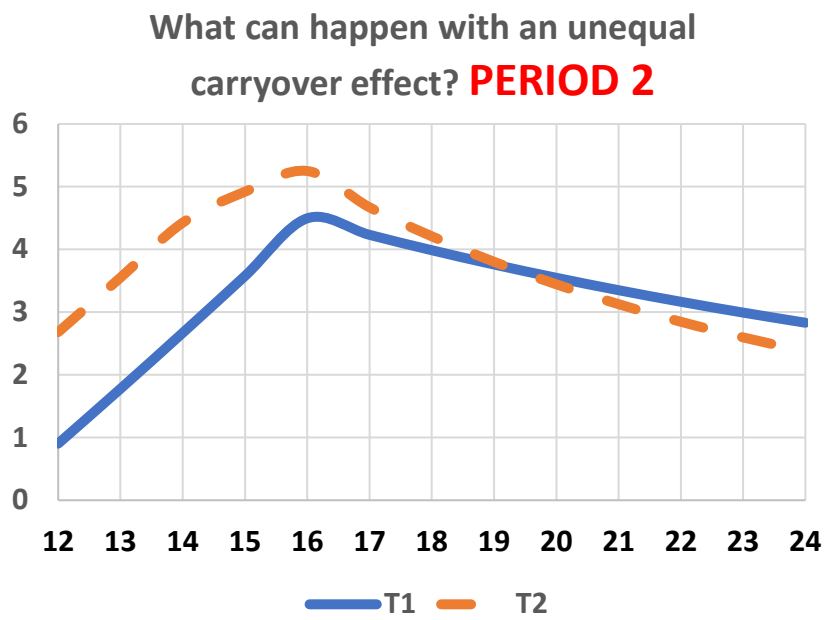
The Importance of Ensuring an Adequate Washout Interval (e.g., an injectable formulation)

A Simulated Example:

- A BE study was conducted using a 2-trt, 2-period, 2-sequence crossover design. However, there was an inadequate washout between periods 1 and 2. Therefore, residual drug from the period 1 treatment contaminated the samples taken for the period 2 treatment.
- In this example, the release of T1 was slower than that of T2, leading to product-specific differences in the magnitude of the drug carryover.
- **Could unequal carryover influence our BE decision?**

The Impact of Inadequate Washout and Unequal Carryover





During period 1, T1>T2. However, during period 2, T2>T1.

Disregarding Per 2 and using only the results from Per 1 would lead to study failure both because of the treatment differences and because of the likelihood of too few observations to provide the necessary study power.

Moreover, the presence of unequal carryover alone is reason to reject a conclusion of product BE.

VICH GL52: Bioequivalence must be tested in each major target animal species on a product label

Due to physiological differences between animal species, we cannot assume that what is bioequivalent in one species will be bioequivalent in another.

Formulation-By-Species Interaction: subcutaneous injection

Relative bioavailability of two **ivermectin formulations** in swine and calves
Lifschitz, et al., (1999). *J. Vet. Pharmacol. Therp.*, 22: 27-34.

Swine

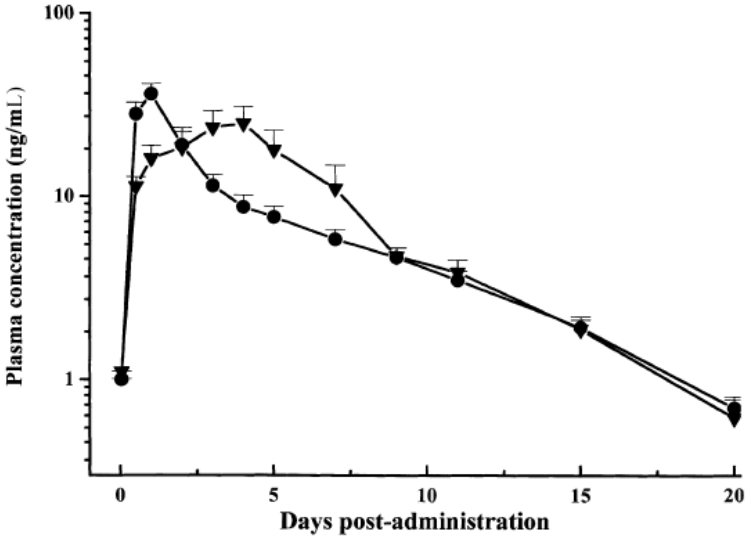


Fig. 1. Mean plasma (± SEM) concentrations of ivermectin (IVM) obtained after subcutaneous administration of the IVM-CONTROL (–●–) and IVM-TEST (–▼–) formulations (300 µg/kg) to pigs.

Calves

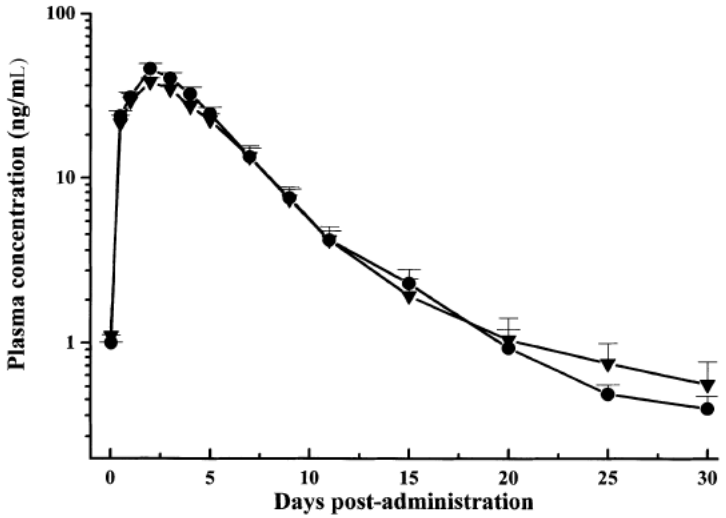
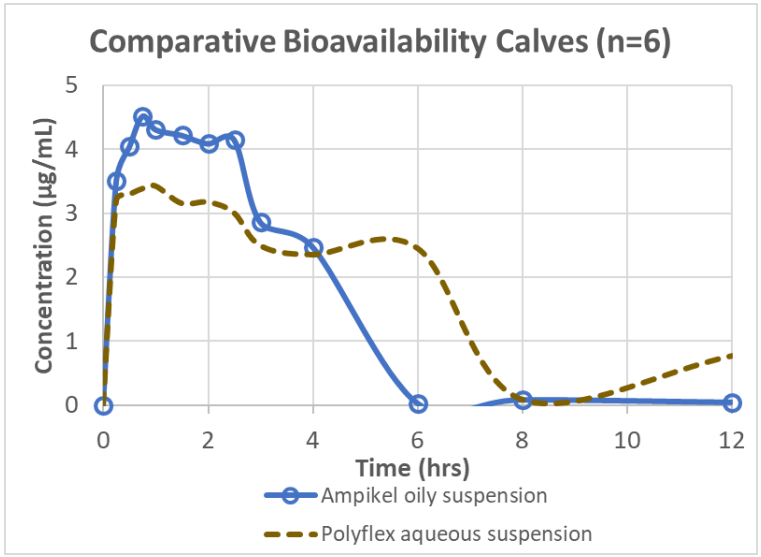
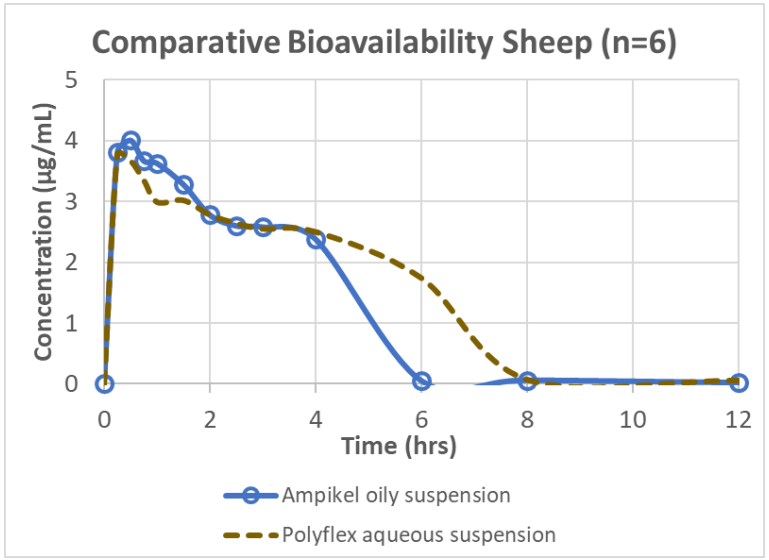
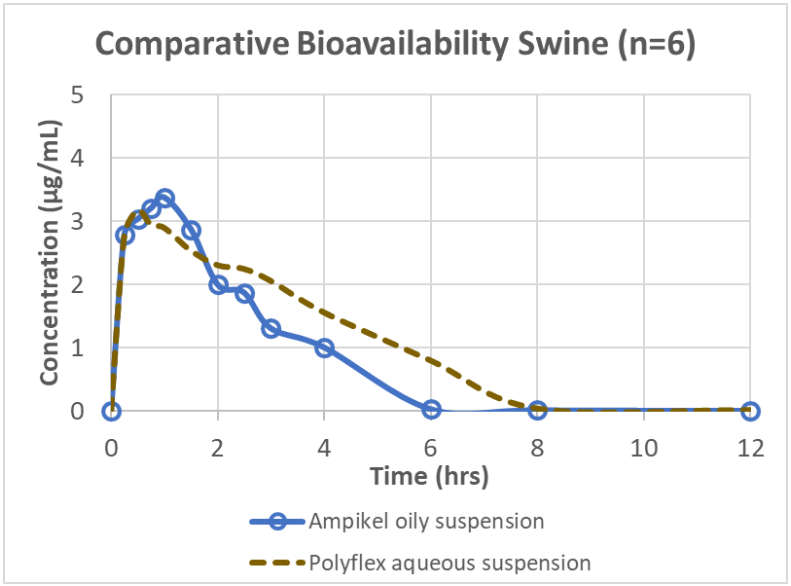


Fig. 2. Mean plasma (± SEM) concentrations of ivermectin (IVM) obtained after subcutaneous administration of the IVM-CONTROL (–●–) and IVM-TEST (–▼–) formulations (200 µg/kg) to calves.

Formulation-By-Species Interaction: intramuscular injection







ampicillin trihydrate aqueous vs oily suspension

Martinez, et al., (2001). *J. Vet. Pharmacol. Therap.*, 24: 1-11.

Formulation-By-Species Interaction: Swine vs Dog tablet bioavailability – immediate release versus erodible matrix tablets.

DOG

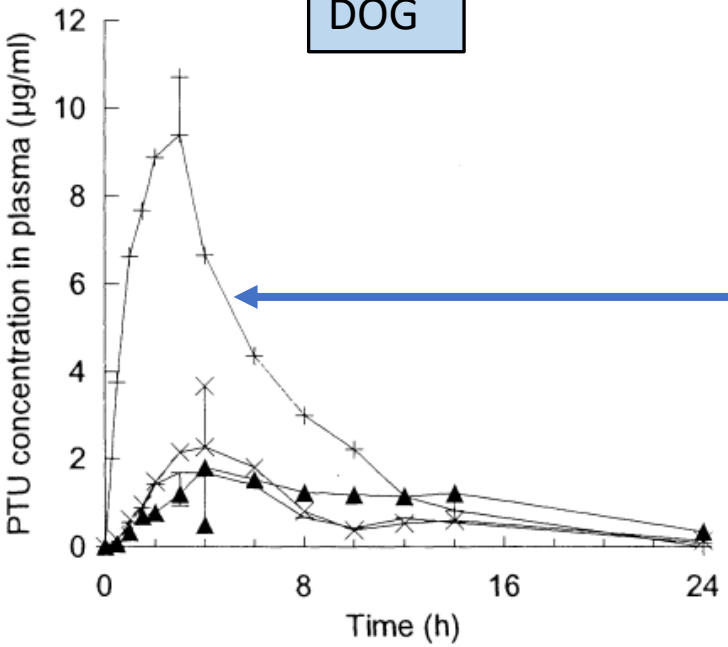


FIGURE 4. Mean concentration-time profiles following oral administration of 300 mg of PTU in dogs. For clarity, the SD is only shown for the peak concentration (n = 6). (+) conventional tablets, (×) K4M matrix tablet, (▲) K15M matrix tablet and (−) K4M:K15M 50/50 matrix tablet.

SWINE

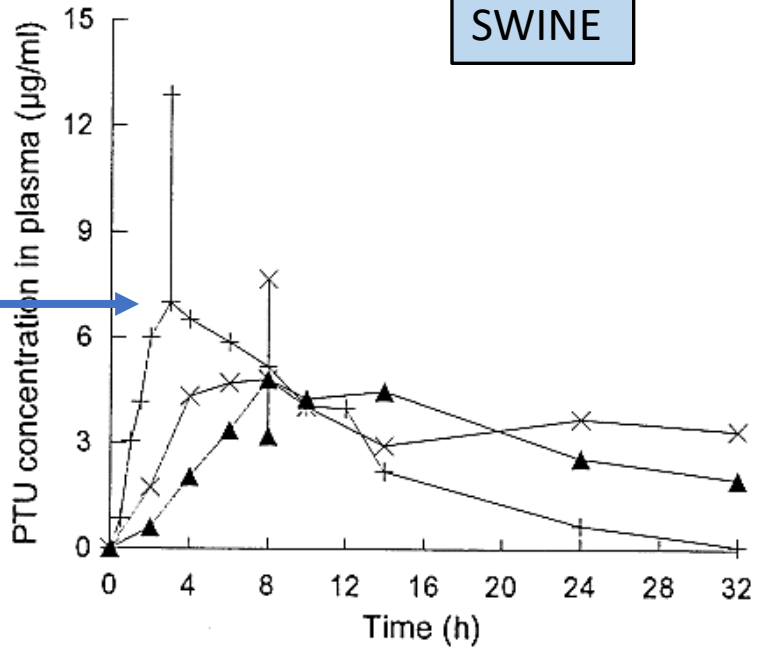


FIGURE 6. Mean concentration-time profiles following oral administration of 300 mg of PTU in pigs. For clarity, the SD is only shown for peak concentration (n = 6). (+) conventional tablets, (×) K4M matrix tablet and (▲) K15M matrix tablet.

Kabanda, *et al*., Pharm Res. 1994 Nov;11(11):1663-8.

Mathe-magics
Adjusting Study Interpretation to the Type of Statistics Used

Comparison Of Various Kinds Of Mean Values: Arithmetic, Harmonic, Geometric

Table 2. Example of differing summary values depending upon the underlying assumptions and estimation procedures for obtaining the mean.

	Number	Ln number	1/Number
	11	2.4	0.09
	7	1.95	0.14
	9	2.20	0.11
	4	1.39	0.25
	10	2.30	0.10
	12	2.48	0.08
	23	3.14	0.04
	15	2.71	0.07
	7	1.95	0.14
	18	2.89	0.06
Type of estimate	**Arithmetic**	**Geometric**	**Harmonic**
Mean	11.60	10.38	9.20
Stdev	5.70	5.29	5.06
%CV	49.14	50.96	50.03

Martinez and Bartholomew, Pharmaceutics. 2017 Apr 13;9(2). pii: E14.

Harmonic mean = the reciprocal of the arithmetic mean of reciprocals.

Geometric mean = the back-transformed mean of the log-transformed variables.

(Sokal and Rohlf, 1973)

When we conduct the statistical analysis of the BE study data, we use geometric means for the comparison of AUC and Cmax. More on this later.

Arithmetic Versus Least Square Means: Influence On Data Interpretation: Example 1

A fictitious study was conducted to answer the question: *did the group that engaged in an exercise routine end up weighing less than the group that did not*?" The parallel study had a different number of males and females per group.

Table 11. Arithmetic means versus LSmeans for the effect of exercise on body weights.

	No exercise group		Exercise group	
	Male	Female	Male	Female
	210	150	200	138
	215	168	192	138
	189	145	176	144
	196	160	202	154
	202	166	210	140
		155	189	
		159	176	
		149	188	
		138	192	
		188		
Marginal means	202	158	192	143
Arithmetic mean	173		174	
LSmean	180		167	

Martinez and Bartholomew, Pharmaceutics. 2017 Apr 13;9(2). pii: E14.

Arithmetic Versus Least Square Means: Influence On Data Interpretation: Example 1

Without correction, one might incorrectly conclude that exercise did not impact body weight.

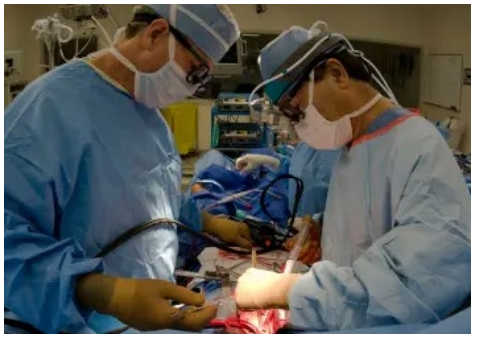
With correction for differences in males and females, it was evident that exercise did lead to a lower BW.

	No exercise group		Exercise group	
	Male	Female	Male	Female
	210	150	200	138
	215	168	192	138
	189	145	176	144
	196	160	202	154
	202	166	210	140
		155	189	
		159	176	
		149	188	
		138	192	
		188		
Marginal means	202	158	192	143
Arithmetic mean		173	174	
LSmean		180	167	

Martinez and Bartholomew, Pharmaceutics. 2017 Apr 13;9(2). pii: E14.

Arithmetic Versus Least Square Means: Influence On Data Interpretation: Example 2

- A fictitious study was conducted to determine whether Hospital A versus B tended to be more conservative in the number of days a patient is retained.
- Patients were categorized by condition. In this example, we will only consider patients enrolled for orthopedic surgery, obstetrics and gynecology, and coronary care.



Looking at the Appropriate Estimates to Answer the Specific Question

	Hospital A	Hospital B
Orthopedic surgery	38 patients	22 patients
Obstetrics and Gyn (OBGYN)	7 patients	36 patients
Coronary Care (CC)	28 patients	12 patients
Arithmetic mean (days, %CV)	6.64 (24%)	6.56 (32%)
LS means	5.86 days	7.37 days

Orthopedic	7.9 days	9.7 days
OBGYN	3.4 days	4.8 days
CC	6.3 days	7.7 days

Based on the arithmetic mean, we would conclude that, **on the average**, patients remained in Hospital A slightly longer than they remained in Hospital B.

However, if we look more carefully, we find that the reason for this slight difference is the inequality of the distribution of patients within each category. In fact, within all categories, Hospital B tended to retain patients for a longer duration than did Hospital A.

	Hospital A	Hospital B
Orthopedic surgery	38 patients	22 patients
Obstetrics and Gyn (OBGYN)	7 patients	36 patients
Coronary Care (CC)	28 patients	12 patients
Arithmetic mean (days, %CV)	6.64 (24%)	6.56 (32%)
LS means	5.86 days	7.37 days

Orthopedic	7.9 days	9.7 days
OBGYN	3.4 days	4.8 days
CC	6.3 days	7.7 days

If our objective was to examine how "conservative" a hospital is regarding the duration of patient stay, we need to consider the **LS means,** which corrected for the imbalance in the number of patients within each category.

	Hospital A	Hospital B
Orthopedic surgery	38 patients	22 patients
Obstetrics and Gyn (OBGYN)	7 patients	36 patients
Coronary Care (CC)	28 patients	12 patients
Arithmetic mean (days, %CV)	6.64 (24%)	6.56 (32%)
LS means	5.86 days	7.37 days

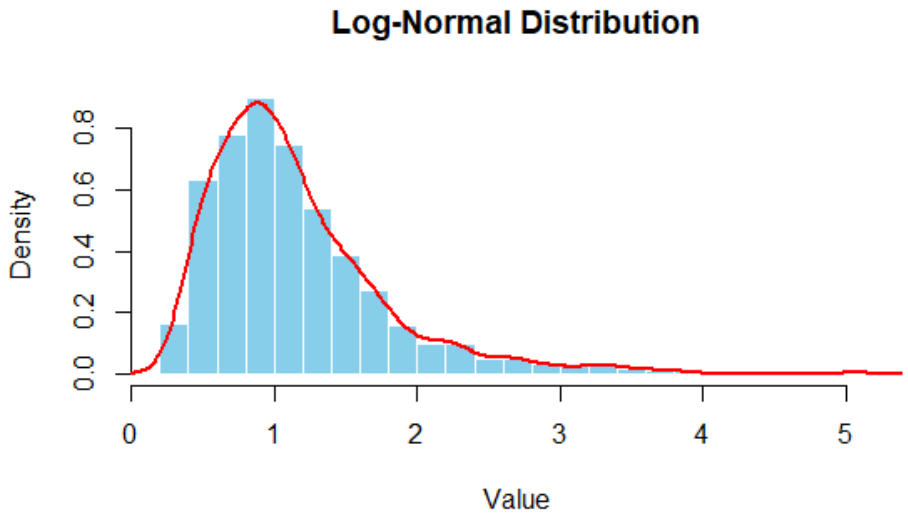
Orthopedic	7.9 days	9.7 days
OBGYN	3.4 days	4.8 days
CC	6.3 days	7.7 days

How Does that Discussion Influence our Evaluation of Product BE?

- When there is no study imbalance (i.e., same number of subjects per sequence and treatment), then the LSmean = the arithmetic mean.
- But means alone are not adequate to generate study inferences about the patient population.
- Not only do we need information on the variability associated with the estimated mean, but we also need a method of evaluating how the study results relate to a potential patient population.

Population Parameter Distribution: Why use Ln-transformation?

AUC and Cmax cannot follow a normal distribution because a normal distribution has positive probability over the entire range of positive and negative values and is symmetric around mean. Rather, the lower bound for AUC or Cmax cannot be less than zero. Therefore, data analysis for product bioequivalence relies on the assumption that both parameters follow a log-normal distribution.



https://sepia2.uni

Population Parameter Distribution: Why Use Ln Transformation?

When evaluating Ln-transformed treatment effects, we generate the treatment comparisons based on differences in the Ln-transformed treatment values. Upon back-transformation of these differences [exponentiation (e^) of differences], the comparison is then expressed in terms of ratios.

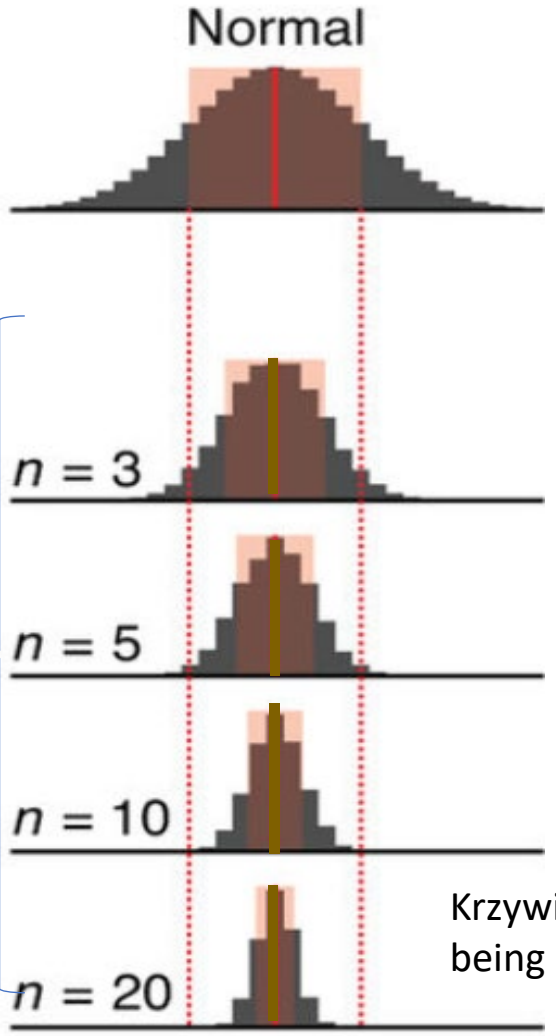
(i.e., e^(Ln A – Ln B) = A/B)

BE assessments are grounded in Test/Reference ratios.

The fundamental difference between **log base 10** (common logarithm) and the **natural log** (**ln**) is their **base**: the common log uses **10** as its base, while the natural log uses the mathematical constant **e** (approximately 2.71828) as its base. Biological processes like drug elimination follow an **exponential function with a base of *e*** (e.g., ke) and therefore using of the natural logarithm simplifies the mathematical operations.

Keep in mind that every study reflects a snapshot of the true population.

Let's say an investigator conducts a 2 period, 2 treatment, 2 sequence crossover BE study. Even if the same investigator repeated the study by sampling from the same population of animals, there would be some magnitude of between-study variability in their results.



Potential spread in study means

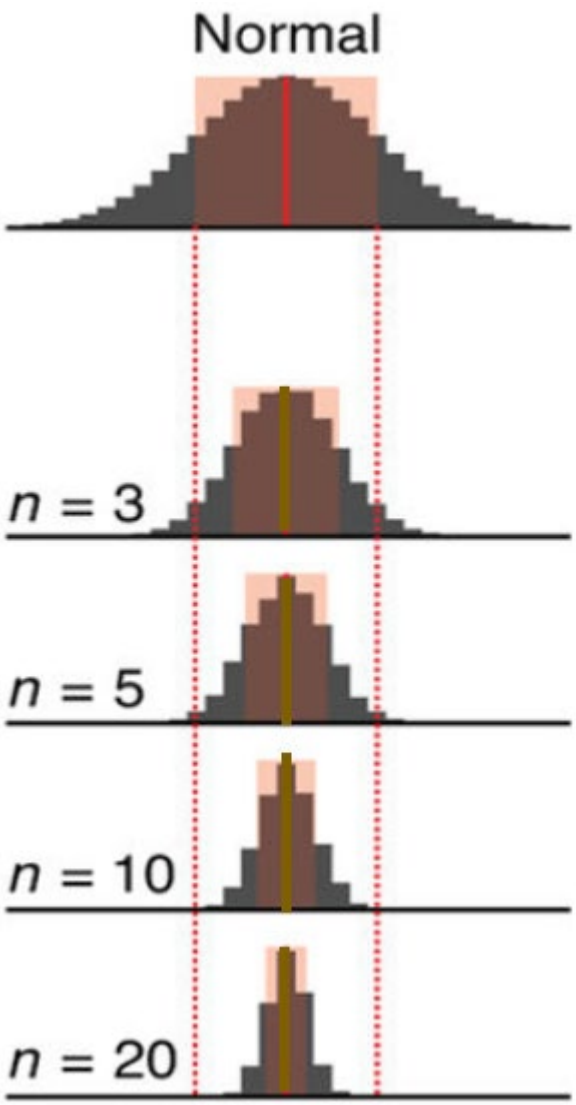
Since we do not know the true population mean, we need statistics that allow for the generation of inferences regarding the true population parameter value from a study with limited sample size.

A very good read!

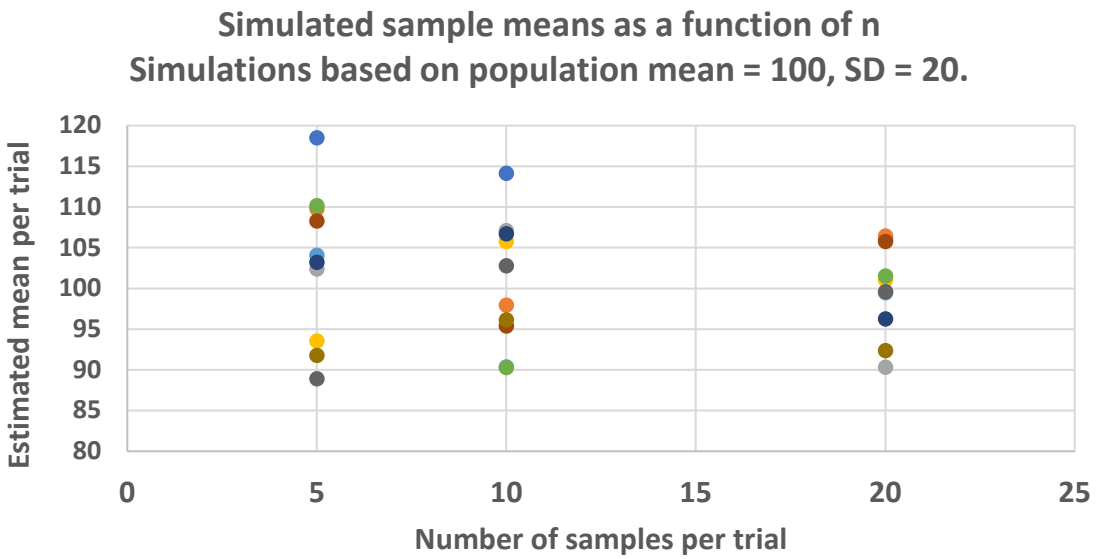
Krzywinski M, Altman N. Points of significance: Importance of being uncertain. Nat Methods. 2013 10:809-10

Note, the spread in the potential sample means are markedly greater with a small versus large sample size. The greater the number of observations, the more precise the potential of sample means. With an infinite number of observations, our study estimate would equal that of the true population mean (the **red** line).

Why is Sample Size Important?



As n increases, uncertainty decreases, (i.e., a more precise estimate). In other words, as n increase, there will be less variability in the potential spread in the estimated means. With greater precision there is a narrowing of the confidence interval.



Setting the Significance Value (α)
Type 1 vs Type 2 Error

H_0 = Patient is pregnant; H_1: Patient is NOT pregnant

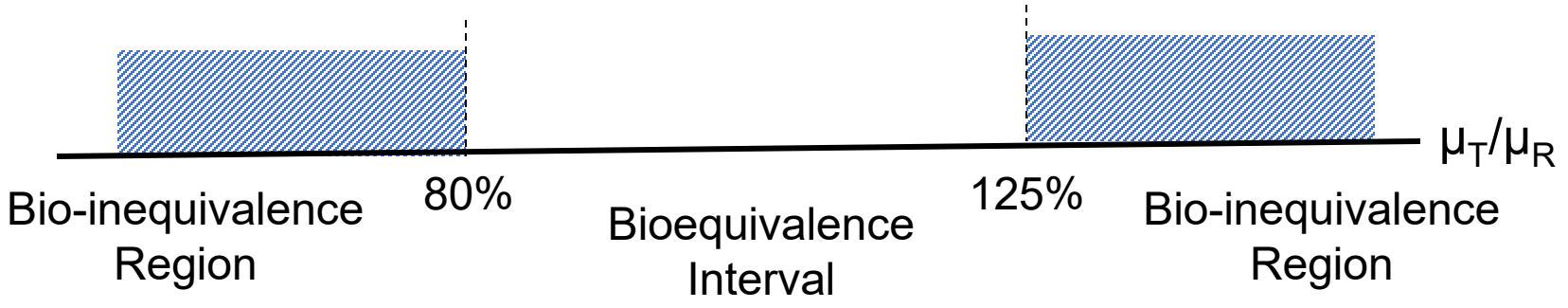
		REALITY	
		H_0 is TRUE	H_0 is NOT True (should accept H_1)
Measured or Perceived Conclusion	Fail to reject H_0	Correct Conclusion. 	Type 2 error (Fail to reject false Ho) 
	Reject H_0 Accept H_1	Type 1 error (Reject a true Ho) 	Correct Conclusion. 

Telling the elderly man that he is pregnant is a Type 2 error (fail to reject the false H_0). Telling the pregnant woman that she is not pregnant is a Type 1 error (incorrectly rejecting the H_0 and accepting H_1).

https://www.youtube.com/watch?v=985KQG-8QV8

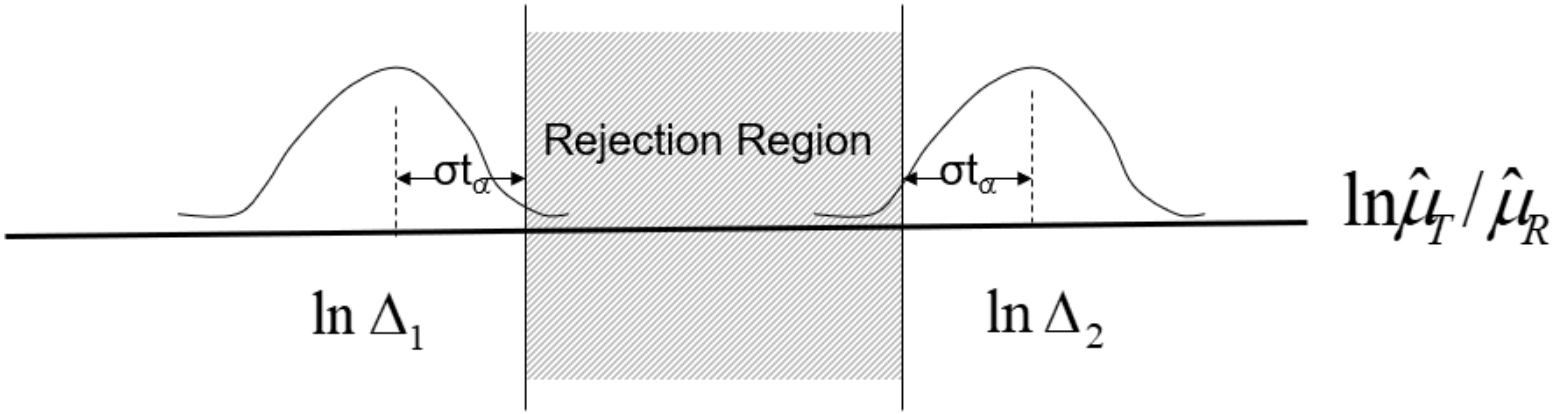
Looking at Type 1 and Type 2 Errors From the Perspective of Product BE.

- In terms of the ratio of the two geometric means μ_T/μ_R, where μ_T and μ_R are the geometric means of the test and reference products, respectively
  - Bioequivalence interval [$\geq$80%, $\leq$125%]
  - Bio-inequivalence regions (0, < 80%) or (> 125%)



When using the 2 one-sided test within the framework of a BE study,

H_0 = bioinequivalence
H_1 = bioequivalence. **Note that we need to reject the H_0 to declare products as bioequivalent.**



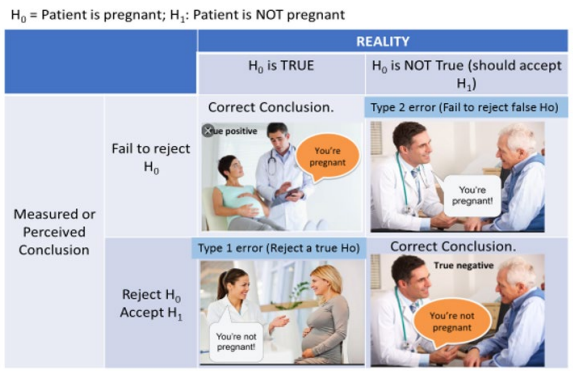
Testing for Bioequivalence: 2 One-Sided Tests

- Hypotheses of testing bioequivalence:
  H_0: $\mu_T/\mu_R < 80\%$ or $\mu_T/\mu_R > 125\%$ vs.
  H_1: $80\% \leq \mu_T/\mu_R \leq 125\%$
- Perform two 1-sided tests
  - H_{01}: $\mu_T/\mu_R < 80\%$ vs. H_{11}: $\mu_T/\mu_R \geq 80\%$
  - H_{02}: $\mu_T/\mu_R > 125\%$ vs. H_{12}: $\mu_T/\mu_R \leq 125\%$
- Reject H_0 if we reject both 1-sided null hypotheses, each at $\alpha=0.05$
- Reject the H_0 (declare equivalence) when a 2-sided 90% CI falling within [80%, 125%]
- The Type I error of rejecting H_0 is controlled at $\alpha=0.05$

Slide developed and previously presented by Donald Schuirmann

Relating Type 1 and Type 2 Error to BE Assessments.

	H_O is true (Products are INEQUIVALENT)	H_I is true (Products are BIOEQUIVALENT
Fail to reject H_O (declare products as inequivalent)	**Correct conclusion**	**Type II error** (bioequivalent products declared to be inequivalent) **SPONSOR RISK**
Accept H_I (declare products as bioequivalent)	**Type I error** (inequivalent products declared to be bioequivalent) **CONSUMER RISK**	**Correct conclusion**



BE Evaluations are Based on Confidence Intervals: What Do They Mean?

- The confidence interval describes the level of **uncertainty** associated with the estimate of interest.
- Confidence intervals are constructed at a specified *confidence level* selected by the user (e.g., for a blood level BE study, this is typically set at 90%, which translates to $\alpha = 0.05$ per tail).

the *NIST/SEMATECH e-Handbook of Statistical Methods* (April, 2012), NIST https://www.itl.nist.gov/div898/handbook/prc/section1/prc14.htm

The 90% Confidence Interval

90% Confidence interval for an estimated parameter *(e.g., mean AUC)*: if the same population is sampled on numerous occasions and interval estimates are made on each occasion, the resulting intervals would bracket the true population parameter (i.e., the T/R ratio) in approximately 90% of the cases.

the *NIST/SEMATECH e-Handbook of Statistical Methods* (April, 2012), NIST https://www.itl.nist.gov/div898/handbook/prc/section1/prc14.htm

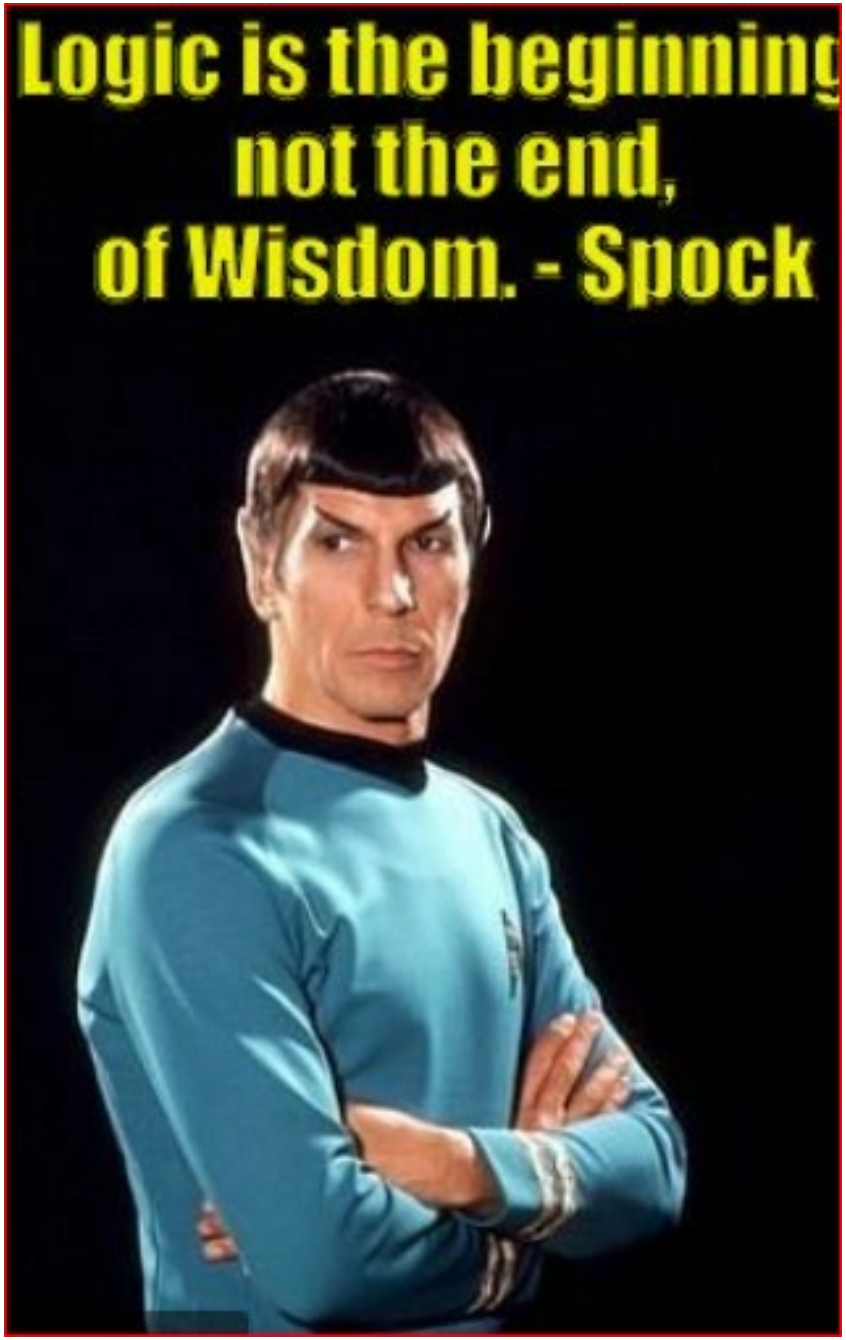
Confidence Intervals: What Do They Mean?

It is appropriate to say that one has 90% **confidence** that the interval contains the true population parameter, but **it is NOT appropriate to say that the interval has 90% probability of containing the true population parameter value.**

Probability quantifies how likely a specific random event is to happen. It's a precise mathematical term (e.g., likelihood of a value when rolling the dice). **Confidence** refers to a level of certainty in a claim or prediction (e.g., previous slide describing the outcome of running a study on same population on numerous occasions).

https://www.nlm.nih.gov/oet/ed/stats/02-950.html

Now for the fun part of this lecture!


Logic is the beginning
not the end,
of Wisdom. - Spock

Data in a crossover study are typically analyzed using an Analysis of Variance (ANOVA).

The ANOVA is used to assess the unexplained variability in the differences between Ln-transformed treatment means.

In a typical 2 treatment, 2 period, 2 sequence crossover study, the subjects are randomly assigned to a sequence.

The assigned sequence determines which treatment is to be administered in the first and second period.

Data in the crossover study are typically analyzed using an Analysis of Variance (ANOVA).

Each subject has values for periods 1 and 2 and for the test and reference treatments.

When evaluating product BE, we need to estimate the unexplained (residual) error (i.e., not explained by our statistical model). This is termed the standard error of the estimate (SE).

Data in a crossover study are typically analyzed using an Analysis of Variance (ANOVA).

Output from an ANOVA provides an estimate of the residual error. This residual error needs to be converted to the Root Mean Square Error (RMSE) = the root of residual error sums of squares divided by the error degrees of freedom. The root mean square error provides a measure of the average model prediction error.

$$\text{RMSE} = \sqrt{\sum_{1=1}^{n} \frac{(\hat{y}_i - y_i)^2}{df}}$$

https://www.kaggle.com/discussions/general/215997

We then convert RMSE to standard error of the estimate (SE)

- To estimate the width of the confidence interval, we need the SE:

- $\mathrm{SE} = RMSE \times \left[\sqrt{\frac{1}{4}\left(\frac{1}{n11} + \frac{1}{n12} + \frac{1}{n21} + \frac{1}{n22}\right)}\right]$

Where n11, n12, n21m and n21 are the number of subjects as a function of period and sequence

Simulated BE Study, 12 subjects per seq

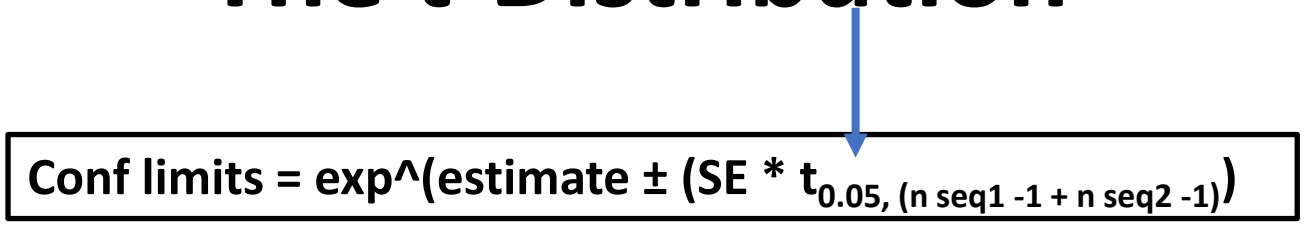
```
data dat;
input subj seq per trt $ auc;
lnauc=log(auc); cards;
1    1    2    R    86.76
2    1    2    R    72.23
3    1    2    R    102.1
4    1    2    R    138.42
5    1    2    R    120.67
6    1    2    R    81.83
7    2    1    R    84.91
8    2    1    R    92.84
9    2    1    R    114.42
10   2    1    R    119.48
11   2    1    R    95.32
12   2    1    R    105.77
1    1    1    T    93.38
2    1    1    T    78.81
3    1    1    T    108.81
4    1    1    T    154.68
5    1    1    T    131.96
6    1    1    T    71.3
7    2    2    T    75.8
8    2    2    T    96.98
9    2    2    T    129.46
10   2    2    T    131.24
11   2    2    T    91.27
12   2    2    T    90.47
```

SAS Code

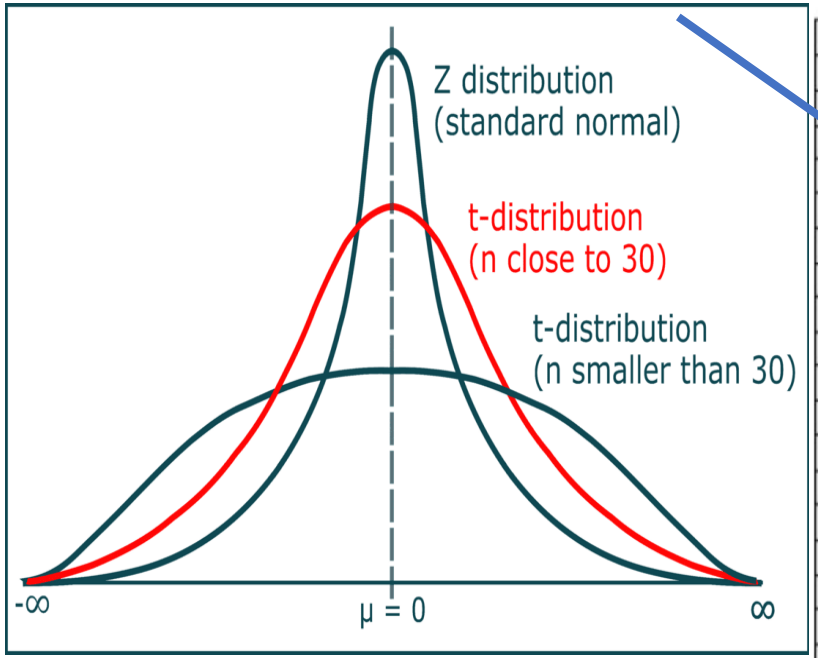
```
;
proc mixed;
classes subj seq  per trt;
model lnauc=seq per trt;
RANDOM subj(seq);
estimate 't - r' trt -1 1/ CL
alpha=0.10;lsmeans trt;run;
```

Note that when run in SAS, the estimate term comparison is generated on the basis of alphabetical order. If using R vs T, then the estimate term = -1 1). If instead, T was called trt 1 and R called trt 2, the estimate term would be written as 1 -1.

The t-Distribution



- Since the estimated difference follows a central t distribution, the corresponding variability in that estimate and a t-value (e.g., 1.81246) are used to calculate the upper and lower confidence bounds.
- The t-distribution is needed for generating confidence bounds because the true population standard deviation is unknown but is estimated from sample data.
- The shape of the t-distribution varies as a function of sample size.



The z-distribution assumes that you know the population standard deviation. The *t*- distribution is based on the sample standard deviation. It is a probability distribution used in statistics when dealing with small sample sizes or when the population standard deviation is unknown.

df/p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869
11	0.259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4.4370
12	0.259033	0.695483	1.356217	1.782288	2.17881	2.68100	3.05454	43178
13	0.258591	0.693829	1.350171	1.770933	2.16037	2.65031	3.01228	4.2208
14	0.258213	0.692417	1.345030	1.761310	2.14479	2.62449	2.97684	4.1405
15	0.257885	0.691197	1.340606	1.753050	2.13145	2.60248	2.94671	4.0728
16	0.257599	0.690132	1.336757	1.745884	2.11991	2.58349	2.92078	4.0150
17	0.257347	0.689195	1.333379	1.739607	2.10982	2.56693	2.89823	3.9651
18	0.257123	0.688364	1.330391	1.734064	2.10092	2.55238	2.87844	3.9216
19	0.256923	0.687621	1.327728	1.729133	2.09302	2.53948	2.86093	3.8834
20	0.256743	0.686954	1.325341	1.724718	2.08596	2.52798	2.84534	3.8495
21	0.256580	0.686352	1.323188	1.720743	2.07961	2.51765	2.83136	3.8193
22	0.256432	0.685805	1.321237	1.717144	2.07387	2.50832	2.81876	3.7921
23	0.256297	0.685306	1.319460	1.713872	2.06866	2.49987	2.80734	3.7676
24	0.256173	0.684850	1.317836	1.710882	2.06390	2.49216	2.79694	3.7454
25	0.256060	0.684430	1.316345	1.708141	2.05954	2.48511	2.78744	3.7251
26	0.255955	0.684043	1.314972	1.705618	2.05553	2.47863	2.77871	3.7066
27	0.255858	0.683685	1.313703	1.703288	2.05183	2.47266	2.77068	3.6896
28	0.255768	0.683353	1.312527	1.701131	2.04841	2.46714	2.76326	3.6739
29	0.255684	0.683044	1.311434	1.699127	2.04523	2.46202	2.75639	3.6594
30	0.255605	0.682756	1.310415	1.697261	2.04227	2.45726	2.75000	3.6460
z	0.253347	0.674490	1.281552	1.644854	1.95996	2.32635	2.57583	3.2905
CI	———	———	80%	90%	95%	98%	99%	99.9%

https://www.geeksforgeeks.org/students-t-distribution-in-statistics/
https://www.youtube.com/watch?v=32CuxWdOlow
https://www.jmp.com/en/statistics-knowledge-portal/t-test/t-distribution#

Estimating degrees of freedom (df) when using a 2 treatment, 2 period, 2 sequence crossover study design (no missing values)

Source of Variation	Degrees of Freedom
Sequence	1
Subject (sequence)	n-2
Period	1
Treatment	1
Residual Error	n-2
Corrected Total	(2n)-1

So, in this example, with 12 subject per sequence, the residual error df = 12-2 = 10.

We then convert RMSE to standard error of the estimate (SE)

Covariance Parameter Estimates	
Cov Parm	Estimate
subj(seq)	0.04832
Residual	0.005366

- Take the square root of 0.005366= 0.073253 = RMSE
- Multiplying that by

$$\left[\sqrt{\frac{1}{4}\left(\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}\right)}\right]= 0.40825$$

SE = 0.073253 *0.40825 = 0.02991

Label	Estimate	Standard Error	DF
t - r	0.01958	0.02991	10

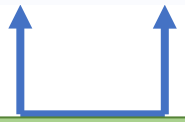
Simulated BE Study Results

Conf limits = exp^(estimate ± (SE * $t_{0.05, (n\ seq1\ -1\ +\ n\ seq2\ -1)}$)

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
seq	1	10	0.00	0.9529
per	1	10	0.89	0.3667
trt	1	10	0.43	0.5274

exp(.01958-(0.02991*1.8125))=exp(-0.03462)= 0.966
exp(.01958+(0.02991*1.8125))=exp(0.07378)= 1.077

Estimates								
Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
t - r	0.01958	0.02991	10	0.65	0.5274	0.1	-0.03462	0.07378



Exp(- 0.03462) = 0.966
Exp(0.07378) = 1.077

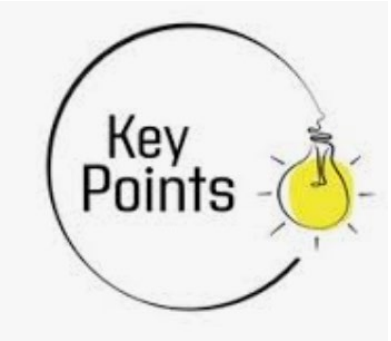
The 90% Confidence Interval = 0.966 – 1.077. Therefore, Test and Ref are BE!

Take Home Exercises

In these exercises, you will have a chance to:

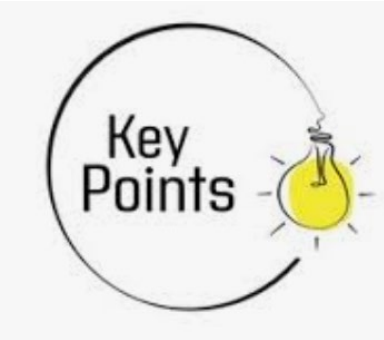
- Case 1: calculate AUC0-last from individual subject data and determine if the test and reference products are bioequivalent.
- Case 2: determine whether a study conducted using a parallel versus crossover design both meet the criteria for BE if the test and reference products have identical PK but where there is substantial variability in drug clearance and absorption.
- Feel free to contact me if you have questions with the analysis

Summary of Key Points



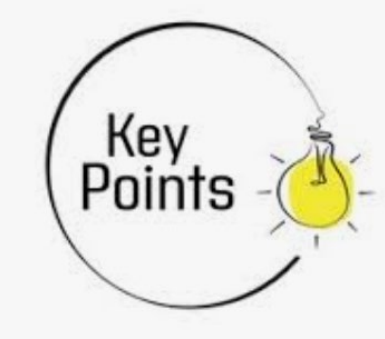
- Linear up-log down method of estimating AUC is NOT the same thing as saying that AUC (and Cmax) follow a log-normal distribution. One has to do with calculating AUC and the other has to do with population predictions.
- Whenever feasible, use of a crossover study design is preferable to a parallel study design.
- Always use the same reference product to avoid drift between generics. Moreover, the potency of the test and reference lots should differ by no more than ± 5%.

Summary of Key Points



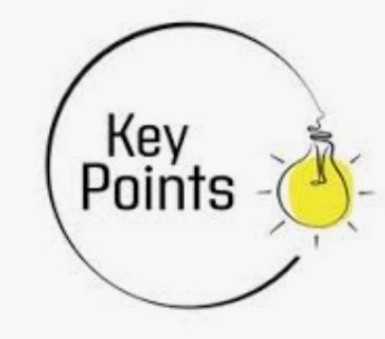
- Unless otherwise specified on the product label, the BE study should be conducted in fasted animals.
- Blood level BE cannot ensure that products will have the same withdrawal period.
- Dose normalization is typically not done in veterinary BE studies.
- It is never appropriate to shave tablets

Summary of Key Points



- Use of the highest label dose only applies to products where there is more than one mg/kg dose approved for use on the label.
- Always ensure that you have an adequate washout interval between study periods.
- A separate BE study is needed for every major animal species on the approved reference product label.

Summary of Key Points



- You should understand the question being asked and consider how the use of means relate to that question.
- When using statistics to assess the results of an in vivo BE study, the hypothesis (H_0) is that the products are not equivalent, and the alternative hypothesis (H_1) is that they are equivalent.
- The upper and lower bounds calculated from the BE study data provides the interval within which we have 90% confidence that it contains the true population parameter.

Thank you all for the opportunity to work with a terrific group of colleagues!

A special **THANK YOU** to Dr. Xin Fang, supervisory statistician, for his review and invaluable recommendations, and to Drs. David Longstaff, team leader, Division of Animal Generic Drugs, and Lisa Troutman, Senior Scientist, Office of New Animal Product Evaluation for their review and suggestions!

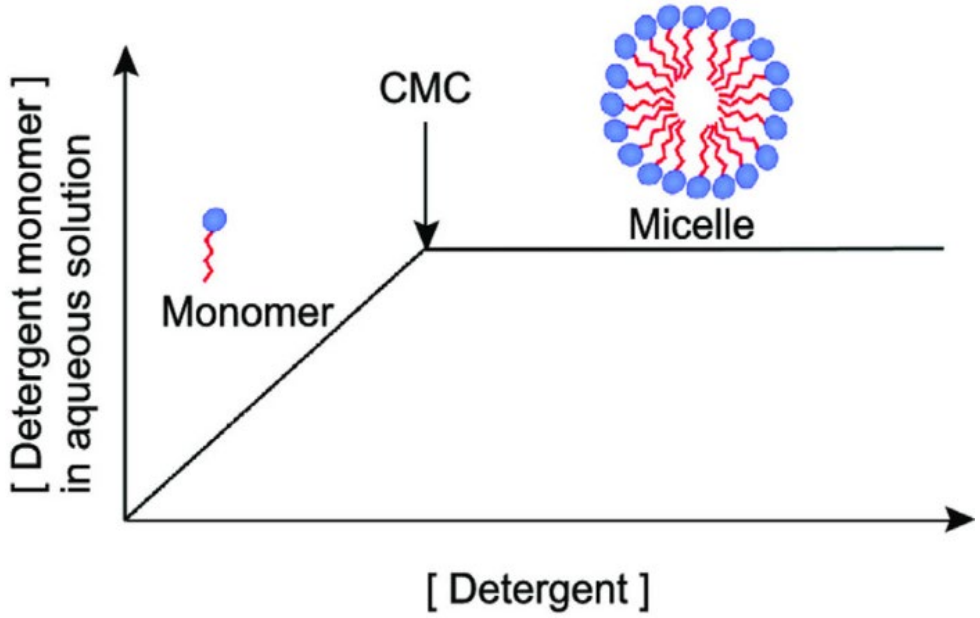
Information about surfactants

Fotaki et al., 2013
surfactants

- The effect of pH on the micellar solubilization of a nonionic surfactant depends solely on the ionization properties of the solubilizate.
- Unionized solubilizates are expected to partition into micelles more favorably than ionized solubilizates.
- The effect of pH on the micellar solubilization by an ionic surfactant will depend on the pKa of the surfactant and the ionization properties of the solubilizate. As the pH decreases towards the pKa of an ionic surfactant, it becomes less soluble, resulting in a lowering of its Critical Micellar Concentration (CMC).
- Generally, the CMC of ionic surfactants is lowered when concentration of dissolved ions is increased.

What is CMC

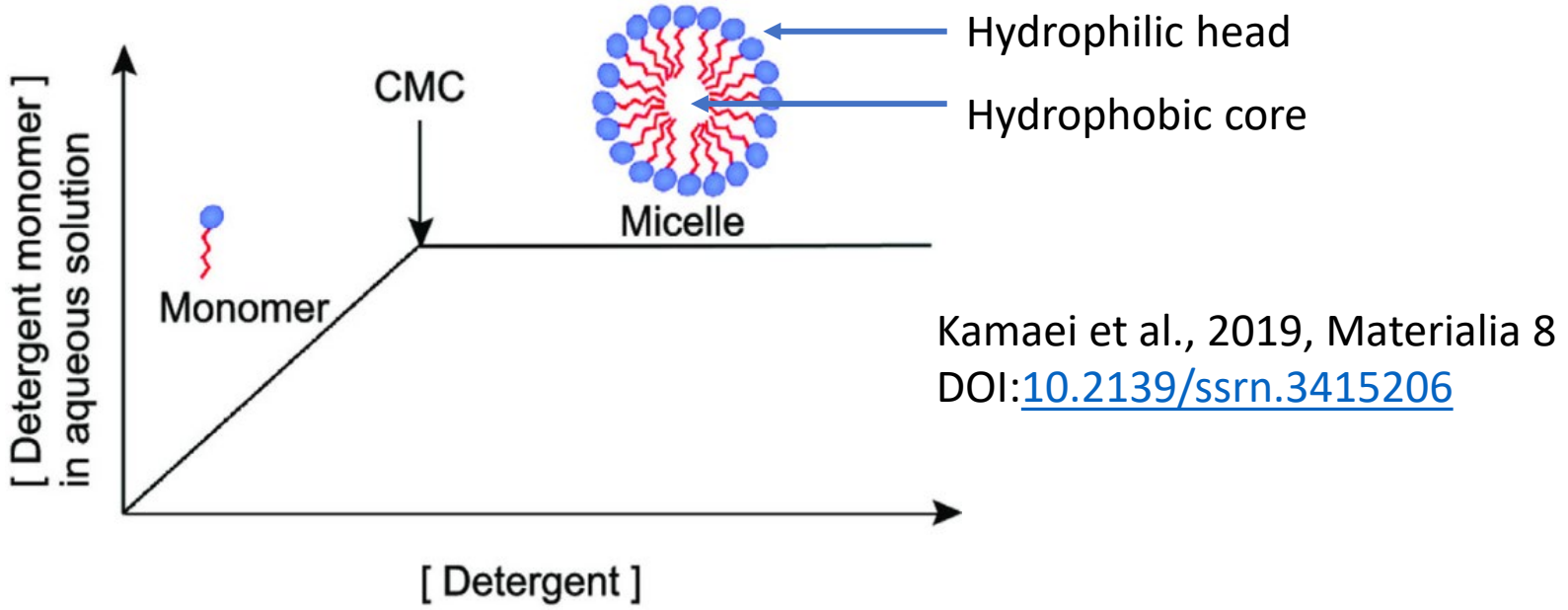
CMC is the concentration of a surfactant in a solution at which micelles begin to spontaneously form . Below the CMC, surfactant molecules exist as individual units (monomers) in the bulk solution or at the surface. Above the CMC, any additional surfactant will self-assemble into micelles to reduce free energy.



Kamaei et al., 2019,
Materialia 8
DOI:10.2139/ssrn.3415206

What is CMC

- **Enhanced solubilization of hydrophobic molecules**: Micelles have hydrophobic cores that can encapsulate poorly water-soluble drug molecules. This greatly increases the apparent solubility of a drug in the aqueous medium.
- **Enhanced dissolution:** The formation of micelles and subsequent solubilization of the drug within them leads to a substantial increase in the drug's overall solubility and dissolution rate.



Kamaei et al., 2019, Materialia 8
DOI:10.2139/ssrn.3415206

Basic –ionized at lower pH.
Expect lower CMC at lower pH

Acidic –ionized at higher pH –
expect lower CMC at higher pH

USP <1092>

Table 1. Commonly Used Surfactants with Critical Micelle Concentrations

Surfactant		CMC (% wt/volume)	Reference
Anionic	Sodium dodecyl sulfate (SDS), Sodium lauryl sulfate (SLS)	0.18%–0.23%	*(2–4)*
	Taurocholic acid sodium salt	0.2%	*(3)*
	Cholic acid sodium salt	0.16%	*(3)*
	Desoxycholic acid sodium salt	0.12%	*(3)*
Cationic	Cetyltrimethyl ammonium bromide (CTAB, Hexadecyltrimethylammonium bromide)	0.033%–0.036% (0.92–1.0 mM)	*(5,6)*
	Benzethonium chloride (Hyamine 1622)	0.18% (4 mM)	*(2)*
Nonionic	Polysorbate 20 (Polyoxyethylene (20) sorbitan monolaurate, Tween 20)	0.07%–0.09%	*(3,7)*
	Polysorbate 80 (Polyoxyethylene (20) sorbitan monooleate, Tween 80)	0.02%–0.08%	*(3,7)*
	Caprylocaproyl polyoxyl-8 glycerides (Labrasol)	0.01%	*(4)*
	Polyoxyl 35 castor oil (Cremophor EL)	0.02%	*(8)*
	Polyoxyethylene 23 lauryl ether (Brij 35)	0.013%	*(9)*
	Octoxinol (Triton X-100)	0.01%–0.03%	*(3,10)*
Zwitterion	Lauryldimethylamine *N*-oxide (LDAO)	0.023%	*(11)*