

## **Session I: Extreme Variability and Aberrant Data in BE Studies (Moderator: Pina D'Angelo)**

April 4, 2019, AM

Speaker #1: Keith Gallicano, Novum Pharmaceutical Research Services

Title: Outliers and Aberrant PK Data in Bioequivalence Studies - Industry Perspective

### Summary:

- Introduced the many different causes of PK outliers.
- Presented examples of different types of within-subject outliers, including those resulting from subject events, study process events (clinic and bioanalytical lab), and a subject-by-formulation interaction.
- Presented a case where excluding a PK outlier led to not concluding bioequivalence; hence the bias in only testing for outliers when it is advantageous to do so; i.e., when excluding a PK outlier converts a failing study to a passing study.
- Proposed setting up a Data Review Board (DRB) to evaluate PK outliers and provided a framework of the DRB, including blinding toward treatment.
- Suggested that good scientific judgment could assign a probable cause, including physiologically implausible drug concentrations, as a reason to exclude unexplained outliers, particularly when there is no documented positive finding from an investigation to support excluding concentrations that are statistical outliers but pharmacokinetically implausible or improbable.

Speaker #2: Dennis Sandell, S5 Consulting

Title: RLDs with High Lot-to-Lot Variability – Issues and a Novel Solution

### Summary:

- Introduced the issue with the reference (RLD) product not being “stable”; i.e., that batches are different and how this impacts BE outcomes.
- Presented data from Bufomix Easyhaler PK studies conducted by Orion and from Advair Diskus 100/50 PK studies conducted by Oriel Therapeutics, concluding that different RLD batches may have significantly different PK; hence, a single RLD batch cannot be selected that represents the product in a reliable manner. The primary reason why PK of RLD batches is different is difference in aerodynamic particle size distribution (APSD) and in particular fine particle dose.
- Questioned whether the current “standard” two-period crossover PK design for showing PK bioequivalence between TEST and RLD products is appropriate.
- Recommended a multi-batch approach to obtain a stable target for the generic development to compare TEST to a RLD “multi-batch”, which is a mixture of samples from several reference batches; noted that this is similar to the approach already used for *in vitro* BE.

Speaker #3: Charlie DiLiberti, Montclair Bioequivalence Services, LLC

Title: When Even Reference Scaling Is Not Enough: Bioequivalence Studies on Extremely Variable Drugs (EVDs)

Summary:

- Introduced that extremely variable drugs (EVDs) with very high intra-subject CVs (e.g., about 80 to 100% or more) are still very challenging to show bioequivalence for, even using the RSABE method.
- Introduced zero PK responses and the bias they cause in practice given that zeros cannot be ln-transformed, and that ln-transformation can greatly exaggerate the importance of T/R ratios (large or small) for low PK responses.
- Found considerable heteroscedasticity across a diverse set of EVD products, particularly among poorly absorbed drugs.
- Presented various possible solutions including a novel approach of using a variance-stabilizing transformation (VST) that can reduce intra-subject CVs dramatically, especially for drugs with highest CVs.
- Noted that VST should 1) be monotonic (preserve order), 2) only transform smaller values, leaving larger values unchanged, 3) address zero values, and 4) be continuous, smooth (i.e., difference between transformed values of zero and transformed values of the smallest possible non-zero values should be very small).
- Noted that VST is not appropriate for addressing zero PK responses for drugs with more modest variability where zero values are not expected and probably reflect failure of dosage unit to release drug, or of the subject to consume it.

Speaker #4: Mark Liu, Mylan Pharmaceuticals, Inc.

Title: Baseline Correction for Endogenous Drugs

Summary:

- Introduced definition and concepts of endogenous drugs.
- Presented statistical concern with current FDA recommendations to set value of negative baseline-corrected plasma concentrations to 0 before calculating the baseline-corrected AUC, as this can introduce biases in the AUC calculation.
- Recommended that time 0 concentration calculated after baseline correction be set to zero instead of the current FDA recommendation to use the calculated baseline-corrected time 0 concentration.
- Questioned if the pre-dose concentration greater than 5% of  $C_{max}$  rule applies to (i) baseline-uncorrected data or (ii) baseline-corrected data, and recommended that the 5% rule does not apply in either case.

Panel Discussion: Robert Lionberger, FDA was an invited panel member.

## Session II: *In Vitro* BE Statistical Issues (Moderator: Sam Raney, FDA)

April 4, 2019, PM

Speaker #1: Elena Rantou, FDA

Title: Statistical Issues with Aberrant IVRT/IVPT Data - FDA Perspective

Summary:

- Presented an overview of the IVPT design and recommended statistical analyses as per the FDA guidance (Acyclovir).
- Power analyses were presented on how to select the optimal number of replicates and donors. Conclusions were that power calculations showed no sensitivity of the power curve to the different estimates of variability. All prior work on pilot study sample size selection indicated a constant improvement in precision, when the sample size increased. Additionally, the choice of the sample size depended on the characteristics and variability of each data set.
- Outlier detection was discussed. Standard practices used in PK-studies (standardized residuals) do not apply because of the small sample size of replicate values within one donor. The question of whether the Dean-Dixon test for outlier detection is appropriate arose for small  $n$  in cases of experimental conduct anomalies that are detected once the sample analysis is completed.
- IVRT was briefly presented. The Wilcoxon Rank-Sum test (as suggested by the SUPAC-SS guidance) was questioned as to whether it is still the most appropriate statistical analysis for IVRT data, in particular: (i) what are the considerations for inflating type-I error under the two-stage structure of the test and (ii) what are the consequences of low power when the coefficient of variation (CV) is high?

Speaker #2: Pina D'Angelo, Novum Pharmaceutical Research Services

Title: Statistical Issues for Low Permeability Compounds in IVPT studies

Summary:

- The statistical issues that arise with zero values (or below the limit of quantitation (BLOQ)) in the dataset were introduced.
- Empirical data were presented followed by suggestions on how to modify the formulae for the point estimate and intra-subject CV.
- Simulations were presented using various intra-subject CVs. T/R Ratios of 100% and 130% were used in the simulations to examine the analyses related to power and type I error, respectively. Three different statistical methods of analyses were presented: (i) RSABE (when  $S_{WR} > 0.294$ ) (ii) ABE using the analyses recommended in the Acyclovir guidance and (iii) ABE using the SAS code recommended in the Progesterone guidance for fully replicate designs.
- Three methods of imputation were examined for BLOQ values: (i) keep as missing, (ii) set to  $\frac{1}{2}$  LOQ and (iii) set to LOQ.

- Imputation of BLOQ values by  $\frac{1}{2}$ ·LOQ generally appeared to overestimate  $S_{WR}$ , which led to a decrease in power especially when  $S_{WR} \leq 0.294$  (ABE) and significantly decreased number of studies that qualify for ABE.
- Treating BLOQ values as missing generally appeared to under estimate  $S_{WR}$ , which led to a decrease in power because of decrease in degrees of freedom, especially when  $S_{WR} > 0.294$  (RSABE).
- Treating BLOQ values as missing generally appeared to increase type-I error ( $S_{WR}$  was under-estimated).
- Results from the ABE analyses ( $S_{WR} \leq 0.294$ ) comparing the statistical method from the Acyclovir FDA draft guidance vs. ANOVA on replicate level data suggested that the FDA draft guidance method may be more powerful in concluding BE; more work to be done on this.

Speaker #3: Diane Potvin, Excelsus Statistics

Title: Sample Size and Statistical Methods Considerations

Summary:

- The objective of this presentation was to evaluate the effect on power and sample size when different numbers of replicates and different statistical models were used.
- Via simulations, results using the mixed-scaled criterion and three methods of statistical analyses were presented: (i) analyses as per the Acyclovir guidance, (ii) a model using a pooled within-donor variance and (iii) a model using separate estimates for the within-donor variance for Test and Reference.
- Various intra-subject CVs were used in the simulations and various numbers of donors and replicates per donor were used.
- Power curves for the results of the simulations were presented.
- The conclusions that were that it was preferable to have more donors with 4 replicates than to have fewer donors with more than four replicates. Further research was also suggested when  $S_{WR} \leq 0.294$  with respect to the model chosen and the impact on power.

Speaker #4: Meng Hu, FDA

Title: Equivalence Criteria for *In Vitro* BE Tests for Locally Acting Drug Products: The Earth Mover's Distance (EMD) Approach

Summary:

- The application of an *in vitro* BE study was presented based on particle size distribution (PSD).
- EMD was presented in the context of population bioequivalence (PBE).
- Method validations using (i) RLD vs. RLD, (ii) RLD vs. negative control and (iii) simulations were presented.
- Conclusions stated that an EMD-based equivalence approach can be used for the complex PSD profile comparison between a generic product and the RLD product. The method

validations showed that the EMD approach was able to effectively reject the unaccepted products (e.g., negative control), and pass the accepted products (e.g., reference itself). The developed approach can potentially be applied for other profile comparison questions for BE purpose.

Panel Discussion: Priyanka Ghosh, FDA and Theo Kapanadze, Diteba were invited panel members.

## **Session III: Practical Issues in BE Statistics (Moderator: Charlie DiLiberti)**

April 5, 2019, AM

Speaker #1: Chuck Bon, Biostudy Solutions

Title: ANOVA Design/Analysis Issues: Nuisance Effects, ANOVA Model Selection, Missing/Unbalanced Data

### Summary:

- In the typical BA/BE study, the only interest is in the treatment effect, so in a way, everything else in the statistical model can be considered to be a nuisance effect.
- In a crossover design, adding subject and period effects increase precision of the model, but adding a sequence effect is a nuisance effect.
- Use PROC Mixed for fully or partially replicated data and Proc GLM for non-replicated data.
- Noted that statistical significance ( $p < 0.05$ ) of a group-by-treatment interaction would be meaningless if the groups were in the same study, same clinic, same population, and similar demographics.
- Noted that statistical significance ( $p < 0.05$ ) of a carryover effect in higher-order crossover designs (3-way or higher) would be meaningless if the treatments contained the same active ingredient(s), the inactive ingredients are not metabolic poisons or inducers, and there are no or low pre-dose concentrations in period 2 or later periods.
- Real life challenges with statistical analysis of an oncology patient PK study with multiple centers and multiple groups were presented.
- Concluded that excluding nuisance parameters (such as sequence effect) could be considered when appropriate.

Speaker #2: Shein-Chung Chow, FDA

Title: Practical Statistical Issues in Evaluation of Average Bioequivalence

### Summary:

- Discussed the differences between the  $1 - 2\alpha$  confidence interval (CI) approach for generic/biosimilar drugs versus the  $1 - \alpha$  CI approach for new drugs. The former is interval

hypothesis testing (two one-sided tests (TOST) at  $\alpha$  level of significance) and is operationally equivalent to  $(1 - 2\alpha)$  CI approach whereas the latter is point hypothesis testing (two-sided test at  $\alpha$  level of significance) and equivalent to  $(1 - \alpha)$  CI approach.

- FDA's recommendation is TOST not  $(1 - 2\alpha)$  CI.
- Emphasized that for a sample size requirement for a standard 2x2 crossover design, in practice, the probability that the constructed 90% CI falls within the BE limits is not the same as the power of establishing BE based on TOST.
- Remarked that power analysis for sample size calculation should be performed using TOST procedure under appropriate study design.
- Remarked that Lund's test for outlier detection is not appropriate and that statistical tests for outlier test should be derived under appropriate study design.
- Commented that missing data imputation may not preserve the overall type I error rate.
- Remarked that there should be a mechanism established for selection of BE criteria depending upon the variability and/or therapeutic index associated with the response of the reference product (e.g., if variability is  $< 10\%$  the BE criterion could be 90-111% for *in vitro* BE testing; if 10-20% the BE criterion could be 85-118% or SABE and if 20-30% then standard 80-125% BE criterion apply for *in vivo* BE testing, and if  $> 30\%$  the BE criterion could be 70-143% or SABE for highly variable drugs).

Speaker #3: Wanjie Sun, FDA

Title: The Effect of Adhesion/Detachment on the Pharmacokinetics of Transdermal Delivery System (TDS)

Summary:

- Evaluated the association between PK parameters and the mean adhesion score in selected ANDA studies for five TDS products, using linear mixed models for the original adhesion data without imputation and adjusting for study design variables.
- Results showed that a higher weighted mean adhesion score (i.e., greater detachment) was significantly associated with a lower level of (log transformed)  $AUC_t$  or  $AUC_{inf}$  (i.e., lower extent of absorption) in four of five TDS products and a lower level of  $C_{max}$  for three of the five TDS products ( $p < 0.05$ ).
- Results for individual plasma concentration versus individual adhesion score indicated that when accounting for the temporal (ADME) effect by plotting plasma concentration vs. adhesion scores at each time point, a clear association between individual plasma concentration and individual adhesion score was revealed (i.e., a higher individual adhesion score (greater detachment) was associated with a lower blood concentration level at most time points, especially at the later ones).
- As more datasets become available, further analysis with paired PK and adhesion results is warranted to verify the apparent trends observed in this work.

Speaker #4: Julie Szirtes, Apotex

Title: PK and Statistical Considerations for Steady State BE Studies – Industry Perspective

Summary:

- Highlighted the key decisions in designing multiple-dose studies, including the challenge of selecting adequate number of doses before PK sampling at steady state for drugs with long half-lives, infrequent dosing intervals, or those having a wide range of half-lives.
- Discussed the importance of adequate washout of drug from period 1 during build-up of drug from switched treatment in period 2 for a crossover design.
- Noted that theoretical differences in attainment of steady state between different periods should not impact the overall T/R ratio on average in a crossover design, but, in practice, differences in half life among individuals will likely not be evenly distributed between the sequences, and, therefore, the T/R ratio may be biased in these situations.
- Presented different statistical tests for evaluating attainment of steady state based on mean and/or individual data, and the challenges associated when the statistical testing does not lead to a conclusion that steady state was achieved.
- Presented challenges for long-acting injectables (LAIs). These included very long half lives leading to flip-flop PK, differences in half-life depending on site of injection and volume of dose injected, wide range of dosing frequency (e.g., every 2 weeks to every 3 months), long study duration leading to high potential for dropouts, bioanalytical stability considerations, susceptibility to PK drug interactions with concomitant medications, and study drug expiry and availability considerations.
- Proposed that other designs should be considered in lieu of the traditional multiple-dose studies for LAIs, such as 1) single dose in healthy subjects, 2) single dose in patients stabilized on other oral therapy, and 3) shortened multiple-dose, switching study in patients stabilized on the study drug.

Speaker #5: Lanyan (Lucy) Fang, FDA

Title: PK and Statistical Considerations for Steady State BE Studies – FDA Perspective

Summary:

- Highlighted the challenges with BE studies of LAIs, including long duration owing to long half-lives, larger sample size owing to high PK variability, and PSGs generally recommend steady-state BE studies in patients owing to safety concerns.
- Presented a case study of paliperidone palmitate suspension injectable (INVEGA SUSTENNA®).
- Current FDA practice for general procedure for steady-state assessment is by comparing at least three pre-dose concentrations for each formulation using linear regression techniques to evaluate achievement of steady state and by comparing the BE outcome with or without the subjects whose terminal slope is significantly different from 0 (i.e., they did not attain steady state).

- Provided an example of an alternative design to the approach used in the PSG whereby a repeated measurements with two consecutive administrations (sequence: TTRR and RRTT) were proposed by the sponsor instead of the conventional steady-state crossover design (sequence: TR and RT).
- Showed via simulations that for a crossover design, products with delayed release will have greater PK differences during the transition period after switch whereas products with slower release will have greater PK differences at new steady state after switch.
- Presented pharmacometric approaches to leverage generic LAI product development, including trial simulations comparing alternative designs and model-based BE assessment.
- Encouraged sponsors to have a pre-ANDA product development meeting to discuss their model and its application with OGD before ANDA submission.

Panel Discussion: Rob Lionberger, FDA and Walter Hauck, Sycamore Consulting were invited panel members.

## **Session IV: Modeling in Bioequivalence (Moderator: Keith Gallicano)**

April 5, 2019, PM

Speaker #1: Murray Ducharme, Learn and Confirm

Title: AUEC and  $E_{max}$  Modeling in Bioequivalence Studies – Industry Perspective

### Summary:

- In regulatory practice (US-FDA) the dose-scale method is recommended for albuterol and levalbuterol MDI products via bronchoconstriction ( $PC_{20}$  endpoint) or bronchodilatation (baseline-adjusted  $FEV_1$  max and  $AUC_{0-6hr}$  endpoints) studies with BE limits (67.00%, 150.00%) and for orlistat via %fecal fat excretion endpoint study (low-dose test and low-dose and high-dose reference; no placebo) with BE limits (80.00%, 125.00%).
- Scientifically, the dose-scale approach should be used when two or more doses can be administered in an equivalence study.
- For albuterol MDI products, the 5-arm design (placebo and low-dose and high-dose test and reference products) rather than the 4-arm design without the high-dose test is preferred to reduce the variability in the  $E_{max}$  modeling.
- Many recommend the bronchoconstriction (i.e., bronchoprovocation) design for bronchodilators, but the bronchodilatation design is simpler and allows the equivalence analysis to be conducted on a clinically useful PD marker (bronchodilatation:  $FEV_1$ ) that is known to follow an  $E_{max}$  relationship.
- Use nonlinear mixed effect (NLME) analysis with either first order (FO: NONMEM, SAS) or first-order conditional estimation (FOCE: NONMEM, more accurate) for the dose-scale analysis when emphasis is theoretically on the population  $E_{max}$  fit and there are missing data (crossover or parallel designs). EM (conditional Expectation/likelihood Maximization) methods (ADAPT5, Phoenix NLME, PPharm, NONMEM) should be used when emphasis is

theoretically on the individual  $E_{max}$  fits and data are available in every subject; use only in crossover designs.

- Results should make sense. If they do not, the log-linear  $E_{max}$  model could be used if data for the reference is between 20-80% of  $E_{max}$ .
- Recommended the need for PD equivalence studies should be dropped from BE requirements, as the lung should not be considered a “topical” organ.
- For skin blanching studies, the simple  $E_{max}$  model appears to be more appropriate versus other more complicated ones. The EM method appears to be more appropriate than FOCE (NONMEM 7) for  $ED_{50}$ . Assumption of In-normal distribution of  $ED_{50}$  should be favored.

Speaker #2: Zhichuan (Matt) Li, FDA

Title: Dose Scale ( $E_{max}$ ) Modeling in Pharmacodynamic BE Studies – FDA Perspective

Summary:

- Summarized the Guidance requirements for dose-scale methodology for one dose level of test (method 1) and two dose levels of test (method 2), and the challenges associated with calculation of the 90% CI for F (relative bioavailability) using repetitive sampling with replacement (bootstrap) to generate “sample dose-response dataset”.
- Discussed three modeling approaches: 1) naïve average data (NAD): mean data only with one data point per dose, 2) naïve pooled data (NPD): treating all data as if they are from the same subject, and 3) nonlinear mixed effect modeling (NLME): all individual data. NLME approach has been routinely used for  $E_{max}$  modeling fitting.
- Tips: Resample by subjects (including data from all treatment arms); use large number of resampled datasets (e.g., 10,000); and use individual data for the test arm in NPD or NLME approach for estimating F by fitting the  $E_{max}$  model to each “sample dose-response dataset”.
- Presented case example of missing  $PC_{20}$  data in bronchoprovocation study, whereby data may be missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR); missing data may impact the F estimation.
- Tip: When there are MNAR, the NLME model is less sensitive to missing values.
- In a bronchoprovocation study, subjects that receive the maximum concentration of methacholine and do not achieve at least a 20% decrease in  $FEV_1$  should be excluded from the statistical analysis with no imputation of the “null” value (i.e., maximum concentration).
- Tip: A bronchoprovocation study may provide more sensitive means of demonstrating BE between test and reference albuterol/levalbuterol MDI product.
- Closing Remarks: Dose-scale modeling is a viable approach to demonstrate PD equivalence for locally-acting drug products with a nonlinear dose-response relationship.

Speaker #3: Liang Zhao, FDA

Title: Use of Modeling and Simulation to Support New BE Approaches

Summary:

- Quantitative methods and modeling (QMM) has been increasingly applied by the FDA to facilitate generic drug development and review and is playing a critical role in the modernization of BE assessment and aiding the development of novel BE methods, *in vitro*-only BE approaches, and risk-based evaluations.
- The scope of QMM activities include regulatory and research activities for quantitative clinical pharmacology models, physiologically-based PK (PBPK) models for systemically and locally-acting products to support not conducting comparative clinical endpoint or PD endpoint studies, and PK/PK-PD based virtual BE studies for study design and alternative pathways.
- Concluded that modeling and simulation has critical impact on generic drug review and approval (see FDA publication in Clin Pharmacol Ther 2019 Feb;105(2):338-349).
- Looking into the future, more collaborations between the Agency and generic industry are key to the successful value creations for generic and new drug development and approval via QMM.

Panel Discussion: Rob Lionberger, FDA was an invited panel member.