Statistical Inputs in Deciding Go/No Go for Generics and Biosimilars in Early Development

Christian Russel Reyes, Ruffy Guilatco and Mary Jane Cadatal

Abstract

Due to high cost of drug development and low likelihood that a drug will reach regulatory approval, an early development plan is needed to help determine the success of the drug development program. This paper will examine the role of a statistician in developing an early high-level clinical development plan for generics and biosimilars with inputs from the structural and functional analysis, and non-clinical results. Discussions will include the statistical tools and techniques used to overcome challenges in developing study design and estimating sample sizes to determine length, size and cost of clinical studies. Finally, the paper will also touch on simulating the probability that the test drug will meet technical success based on pilot studies or published data from the originator drug.

Keywords: clinical statistics, generics, biosimilars, bioequivalence, probability of success.

Introduction

Developing a new drug would take at least US$1 Billion. And only 1 out of 10 new drugs will reach regulatory approval. Development of a generic on the other hand would take about US$10 Million with a success rate in reaching market authorization from phase 1 of at least 90%. Biosimilars development would take US$50 Million to US$100 Million due to complexity of the drug with a probability of success of around 78%. Generics costs about 80% to 90% cheaper than originator drug while biosimilars costs around 20% to 30% cheaper. A good drug characterization from pre-clinical analysis including structural and functional analysis and a strategic clinical plan are needed to develop a cost-effective alternative. This paper will investigate the challenges usually encountered by statisticians in developing the plan and the techniques and tools used to overcome some of these challenges.

Generics and Biosimilars

US Food and Drug Administration (FDA) has defined a generic drug as a medication created to be the same as an already marketed brand-name drug in dosage form, safety, strength, route of administration, quality, performance characteristics, and intended use. These similarities help to demonstrate bioequivalence, which means that a generic medicine works in the same way and provides the same clinical benefit as its brand-name version. While a biosimilar is a biological product that is highly similar to and has no clinically meaningful differences from an existing FDA-approved reference product. The development of generics and biosimilars is very different from that of a new drug. Barbier, Declerck, et al., (2019) summarizes the types of information needed in drug development. The bulk of data needed to show effectiveness and safety for a new medicine comes from clinical trials. While in biosimilar development and similarly for generics, most of the data comes from analytical and functional analysis, and from preclinical studies. In most cases where the generic medicine was shown to have no meaningful difference from the originator drug based on these data, a biowaiver can be applied even without conducting clinical trials. Once approved, all indications of the originator drug apply to the generic or biosimilar product.

Bioavailability (BA) and Bioequivalence (BE)

The US FDA defines bioavailability as the rate and extent to which the active ingredient or active moiety is absorbed from a drug product and becomes available at the site of drug action. For drug products that are not intended to be absorbed into the bloodstream, bioavailability may be assessed by scientifically valid measurements intended to reflect the rate and extent to which the active ingredient or active moiety becomes available at the site of drug action. Bioequivalence on the other hand is the absence of a significant difference in the bioavailability between reference drug and in pharmaceutical equivalents or pharmaceutical alternatives when administered at the same molar dose under similar conditions in an appropriately designed study. BE comparisons normally rely on (1) a criterion, (2) a confidence interval (CI) for the criterion, and (3) a predetermined BE limit. BE comparisons could also be used in certain pharmaceutical product line extensions, such as additional strengths, new dosage forms (e.g., changes from immediate release to extended release), and new routes of administration. In these settings, the approaches described in this guidance can be used to determine BE. Selected pharmacokinetic (PK) parameters and preset acceptance limits allow the final decision on bioequivalence of the tested products. The area under the concentration time curve (AUC) is often used to measure the extent of absorption or total amount of drug absorbed in the body. The maximum plasma concentration or peak exposure (CMAX), and the time to maximum plasma concentration (TMAX) are parameters that are influenced by absorption rate. Other parameters that are usually investigated are Clearance (CL), Elimination Half-life (t1/2), and Elimination Rate Constant (ke).

A good discussion on the analysis of bioequivalence can be found in Chow and Liu (2009). They note that regulations in most countries including United States, EU and Japan require only that evidence of average bioavailability be provided for approval of a generic drug. For biosimilars clinical trials, PK/PD and clinical immunogenicity may be required. In this paper, we will focus on statistician's role in providing inputs in designing a plan for bioequivalence studies.

Challenges

Below are some of the challenges encountered when designing a clinical trial for a generic or a biosimilar.

1. No regulatory guidance and limited references on the originator drug
   One of the challenges in designing a clinical trial with the objective of providing evidence of bioequivalence between the test drug and the reference drug is collating information on the reference drug. Regulatory agencies require pharmaceutical agencies to divulge information relating to the safety and efficacy of their drug. However, it is often a case that some of the information are redacted or information on the bioavailability of the drug is insufficient for designing a bioequivalence study.

2. Designing for multiple markets/countries
   One consideration in creating a high-level plan is whether a single clinical trial will be sufficient for submissions in different markets. There will be a need to check the regulatory guidance of each country. There is also a need to determine whether the originator drug has a different formulation in each market. In cases where there are substantial differences between reference drugs in two different markets, there will be a need to design two separate studies for each formulation or a single study comparing the test drug to both formulations.

3. Drugs with long half-life
   In designing a crossover study, one concern is a possible carryover effect, wherein the drug administered in one period affects the results in the next period. A washout period

is usually included to control carryover effect. But this will be difficult for drugs with long half-life as this will affect total duration of study and possibly increase patient dropout rate.

4. Drug Toxicity
   Drug Toxicity should also be considered in the design of the study. In most cases, regulatory agency will recommend the most sensitive population, which is usually healthy volunteers. For cytotoxic drugs, an in-patient study might be more appropriate. The choice of population may influence the length of the study, drop-out rate and variability of the measures.

5. Highly Variable Drugs
   Another consideration in designing a PK/PD study is when the reference drug is highly variable. US FDA and European Medicines Agency (EMA) consider a reference drug as highly variable if the within-subject coefficient of variation is higher than 30%. Bioequivalence will be difficult to demonstrate unless the sample size is very large.

6. First in Market Biosimilar
   Regulatory agencies may require safety and efficacy studies for first in market biosimilar. The approval may also go through an advisory committee to the regulatory agency.

Other considerations in planning a clinical trial may include the use of sensitive populations for in-patient studies, cost of doing a clinical trial for a certain market, and compliance to regulatory guidance. To overcome some of these challenges from a statistical point-of-view, the following techniques are used.

Statistical theories and techniques

The following are techniques and theories used to work around the challenges or at least come up with estimates that will help in developing the clinical plan.

1. Study Design
   a. Pilot Study/ Sequential Study Design
      In cases where there is very limited information on the bioavailability of the reference drug, a pilot study or sequential study design can be used. Pilot study is a study that can be used to validate analytical methodology, assess variability, optimize sample collection time intervals, and provide other information as described in the US FDA guidance (2003). Pilot study can be conducted in less than 12 subjects but there are inherent limitations doing this as the estimates may be unreliable for sample size planning. According to the US FDA guidance (2011), a minimum of 12 evaluable subjects should be included in any BE study. One advantage of conducting a pilot study is that BE may be claimed in the pilot study if stated as a procedure in the protocol and hence further study will not be necessary.

      In 2007 and 2010, EMA and US FDA published their guidance on adaptive clinical trial designs, respectively. Sequential design is an adaptive design that allows modifications to the trial's design of statistical procedures during the study conduct without undermining the validity and integrity of the trial. In addition, it allows for premature termination of the trial if there are futility, efficacy or safety issues based on the interim results.

      As discussed in detail by Pocock (1977) and O'Brien and Fleming's (1979), the main aim of the sequential design is to decrease the sample size. Interim analysis provides

a chance to make appropriate decisions with regards to the allocation of resources for the clinical drug development.

The methods by Potvin, D., et al. (2008) are the first validated frameworks of two stage sequential design in the context of BE wherein first initial group of subjects are treated and data are analyzed. One advantage of this approach is that in stage one, the trial can be stopped, and BE can be claimed. Otherwise, if bioequivalence is not demonstrated, additional subjects can be added and the results from both groups combined in final statistical analyses.

Generally, in a group sequential trial, interim analyses are conducted on the available data at one or more intermediate stages, as determined by a *priori* decision rules to guide the adaptations. The flexibility of this design reduces the cost and increases the probability of success and time to completion of the clinical development plan. This can also translate to more ethical treatment of patients since there is a possibility of inclusion of fewer patients and hence avoiding unnecessary exposure of many subjects to the drug, and more efficient drug development. The most common adaptive trials used include standard group sequential design, sample size re-estimation, and drop-the-loser design (Jones, B. & Kenward, M.G., 2003 and Chow, S.C. & Pong, A., 2011).

There are advantages using the adaptive design framework, and so do risks and challenges exist. Hence, adequate and careful planning is critical for the success of the trial.

b. Replicate Design

In a replicate design, a subject is randomly assigned to a sequence with at least 1 treatment administered twice. This design is also suggested for highly variable drugs to determine not just global intra-subject variability but also treatment intra-subject variability. For biosimilars, a replicate design can also be used to set reference standards when there is limited public information on the reference drug. A full replicate design is where the subjects either receive the reference or test drug in two periods (ex: RTR, TRT) or all subjects receive the reference or test drug in two periods (ex: RTRT, TRTR). A partial replicate design is a 2-treatment, 3-sequence, 3-period design where the subject will receive the reference drug in two periods (TRR, RTR, RRT).

2. Considerations for Highly Variable Drugs

a. Evaluation of parameters

Haidar S., et al. (2008) discussed the different approaches of regulatory agencies for highly variable drugs. EMA recommends the use of expanding limits (ABEL) for CMAX given that the point estimate is within 80-125%. Canada health applies the usual 80-125% limit for the 90% CI of AUC and for the point estimate of CMAX. US FDA proposes a scaled average bioequivalence (SABE) for both AUC and CMAX. Laszlo and Lazlo (Tothfalusi and Endrenyi, 2012) detailed the steps for the EMA and FDA approaches.

b. Sample Size considerations

High variability would mean larger sample sizes. Crossover designs may be considered in designing a study provided that the drugs have short half-lives. As discussed above, replicate design should be applied when using ABEL or SABE. Also,

larger absolute differences between the two logarithmic means can be noted in the various BE studies when the within-subject variation is higher. Therefore, it is recommended that a 10% deviation between the means, i.e. a true GMR = 1.10, be considered during sample size determination.

3. Pooled Variability and Upper Confidence Limit

One of the most important considerations in designing a bioequivalence study is the sample size calculation and its associated power. Hence to ensure adequate power, finding the optimal sample size is very important. Sample sizes that are too small increase the type II error and may result in study failure; whereas sample sizes that are too large increase the cost of the study and unnecessarily exposing many subjects to the drug. According to the statistical guidelines of the US FDA and EMA, 80% or 90% power is recommended for bioequivalence studies.

In sample size determination, information on the intra-subject coefficient of variation (intra-CV) is needed for pharmacokinetic parameters since crossover designs are often utilized in bioequivalence (BE) studies. However, when different estimates for intra-CVs are produced in various studies with identical generic drugs, choosing the appropriate intra-CV is a challenge. This is the case where pooled CV will be beneficial to have an appropriate estimate of CV.

Patterson and Jones (2017) discussed a method of pooling data across studies. Prior to pooling, variances must be weighted according to the studies' sample size and sequences. Larger studies tend to be more influential than smaller ones, and more sequences (with the same sample sizes) give higher CV. Of note, additivity of variances can be applied in the parametric model of log-transformed data.

Steps in calculating pooled CV%:

a) Calculate the variance from CV

$$\sigma^2 = \ln\left(CV_{intra}^2 + 1\right)$$

b) Calculate the total variance weighted by degrees of freedom (df)

$$\sum \sigma_w^2 df$$

c) Calculate the pooled CV from the total variance

$$CV = \sqrt{e^{\sum \sigma_w^2 df / \sum df} - 1}$$

d) Optionally calculate an upper (1-α) % confidence limit on the pooled CV (recommended α = 0.25)

$$CL_{CV} = \sqrt{e^{\sum \sigma_w^2 df / \sum \chi_\alpha^2 \sum df} - 1}$$

In cases where reliable estimates of the variability from relevant historical data to be used in estimating the sample size is not available, the common practice is to conduct a pilot study to generate an initial estimate of the effect size and intra-CV. Since pilot studies are usually conducted in small number of subjects, there is still uncertainty on the estimated CV, therefore, one conservative approach is to use the upper confidence limit of the estimated CV. Gould (1995) suggested to use the 75% upper confidence limit of the CV in determining the sample size of the main study.

4. Probability of success

Conducting a pivotal clinical trial requires a considerable amount of the company's resources. Therefore, it is important to have an informed decision on whether to proceed

or stop with conducting additional clinical trials early on during the drug development program. One of the key information that could drive the decision-making is the probability of success of a planned clinical trial. The probability of success considers the uncertainty around the estimates of the hypothesis parameter. Probability of success is sometimes called probability of study success, predictive probability of success, expected power, average success probability and unconditional power. The most common approach is to assume a prior distribution on the hypothesis parameter based on available prior data and to take the average or expected value of the power. Spiegelhalter, et al (2004) discussed the difference of classical power and classical power averaged on the prior distribution of the hypothesis parameter. Liu, et al (2010) further developed expected power by incorporating uncertainty on the variance. In case of a bioequivalence test, the uncertainty can come from the geometric mean ratio or on the variability of the parameter. A useful function in the software R-language exppower.TOST gives the user the capability to determine the expected power by specifying whether the uncertainty comes from the GMR, the variability or both.

Applications

Below are some examples of study designs for a generic and for a biosimilar.

Example 1

The first example is an application of the pooled variability, upper confidence limit and highly variable techniques discussed above on a possible generic drug. Data is from two pilot studies with 2-way crossover design on the bioavailability of the drug with only 6 healthy volunteers per study. Mean square errors were computed below from the confidence limits. Using the formulas for pooled variability, pooled CVs are 34.37% and 23.08% for AUC and CMAX, respectively. Given the small sample size, a 75% upper confidence limit is applied to the pooled variance giving 43.9% and 29.21% for AUC and CMAX, respectively. As the variability of the AUC is higher and assuming a 5% level of significance, 90% power and a geometric mean ratio of 5% from 1, a possible design is a 2-way crossover study with 106 evaluable subjects. But given the high variability (CV > 30%), a possible more appropriate design is partial replicate 2-treatment, 3-sequence, 3-period design (RRT-RTR-TRR). Given the same assumptions but with larger range of geometric mean ratio of 10% from 1, it will only take 50 evaluable subjects to reach a power of 90%.

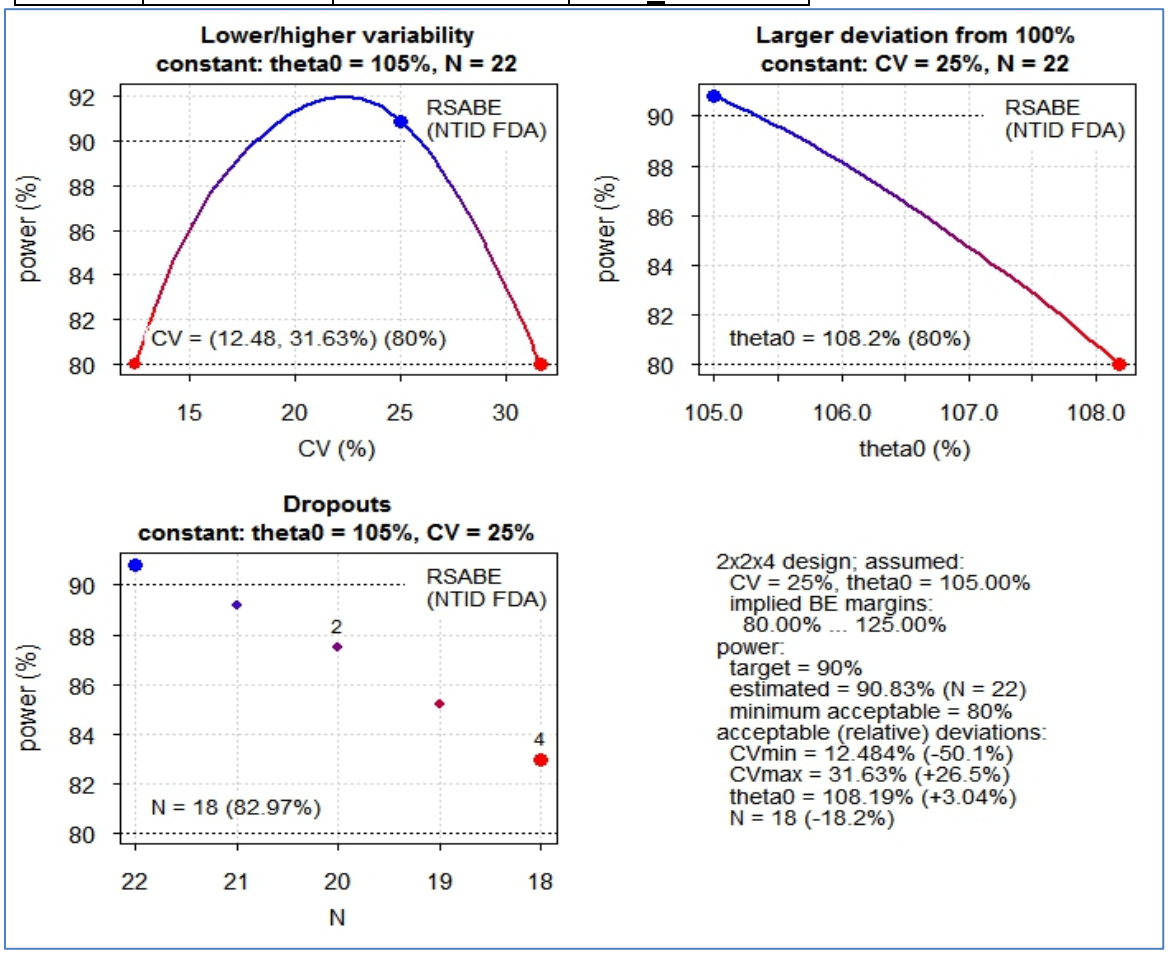
	PK Parameter	MSE	CVintra	N
Country 1	AUC0-t	3.07	35.0%	6
	Cmax	2.89	25.0%	6
Country 2	AUC0-t	3.03	33.0%	6
	Cmax	2.84	21.0%	6

Example 2

A single-dose, randomized, two-treatment, two-period crossover design was conducted between the test and reference narrow therapeutic index drugs (NTID). The objective of the study was to establish bioequivalence for all four analytes (A1, A2, A3 and A4). The CIs of the GMRs of PK parameters for A2, were not contained within the pre-specified BE limit, therefore the study failed to show bioequivalence. After several years, the FDA released a new guidance pertaining to BE demonstration specific to the active ingredient of the NTID, with a design that is totally different to the previously conducted trial, that is, a single-dose, four-way, fully replicated crossover design only on A3 was recommended. The statistician conducted simulation studies using the estimates of GMR and intra-CV of PK parameters for A3 obtained from the initial study. Results showed that if a clinical trial based on the new design recommended by FDA is conducted, the probability of

demonstrating bioequivalence is high, for all simulation scenarios considered. A sample of graphs are shown below to demonstrate that the design has at least 80% power across ranges of values for sample size, GMR or CV.

N	GMR	CV	Power
22	92 to 108%	13 to 32%	≥ 80%
28	92 to 108%	10 to 36%	≥ 80%
36	92 to 108%	9 to 42%	≥ 80%



Example 3

In one biosimilar PK/PD study conducted by Hospira/Pfizer, the formulation of the reference product in the US is different from that in the EU. The study design used was a 3-treatment, 3-period, 6-sequence design with a step-down or hierarchical analysis of PK and PD parameters. In the step-down analysis, the test drug is first compared to the US reference drug. If test result is positive, the test drug is then compared to the EU reference drug. This type of analysis is used to control type I error.

Example 4

Liu (2010) presented an application of Probability of Success on BE studies to be conducted on four formulations. Results of the simulations suggest that across the four formulations considered, one formulation attained the highest estimated probability of success equal to 76% (next highest

had 46%) which could help decision-makers on identifying the formulation that will be included in further studies under the drug development program.

Discussions

As discussed in the introduction, generic or biosimilar drug development is done in stages starting on the structural and functional analysis, then moving to nonclinical and clinical studies. Even prior to conclusion of non-clinical studies, high-level clinical plans are developed to determine the final cost of the drug development and the length of time to regulatory submission and market release. Overall drug development plan across the different stages will help management decide whether to go through with the drug development or not (Go/No Go decision). A too conservative clinical plan will lead to higher study cost, longer length of the study and higher risk to patients. This plan can potentially lead to a No-Go decision. A too optimistic plan may lead to an underpowered study and potential failure of the study. Study failure will increase the cost and delay the development of the drug. The statistician plays a critical role in developing these clinical plans working with a cross functional team of clinicians, pharmacologists, clinical managers and regulatory experts among others.

Statisticians will need to review prior information on the reference drug and on the test drug to determine the most appropriate study designs and sample sizes. Results from animal studies can be used as guidance on the possible toxicity of the drug. And in some cases, human pharmacokinetic profiles are simulated using animal data (Mordenti, 1985) to get an idea on the drug's bioavailability. But in most cases, a clinical bioequivalence study is required to get approval if a biowaiver is not possible. The statistician will also need to consider the uncertainties of the estimates from prior information along with other variabilities. Some of the techniques and tools that can help the statistician assess these uncertainties to come up with a more robust study design are discussed in this paper. Though the discussions in this paper are not comprehensive, the main objective is to give statisticians awareness on the usual challenges encountered and possible workarounds. In coming up with most appropriate study design, the statistician needs to work with the pharmacologist and clinician to consider drug characteristics and toxicity. The inputs for study design and sample size are then used by the project manager to come up with an estimate on the development cost and target release in the market.

Finally, the main goal for developing biosimilars and generics is to provide patients with safe, effective and cheaper alternatives to drugs in the market. By developing an accurate, robust and efficient clinical plan, the statistician along with the whole functional team, can help the company reach this goal.

Acknowledgments

This work was supported by Pfizer Inc. The authors would also like to thank Josephine Domingo and Junibel Ramos for their reviews and inputs.

References:

1. Barbier, L., Declerck, P., Simoens, S., Neven, P., Vulto, A. and Huys, I. The arrival of biosimilar monoclonal antibodies in oncology: clinical studies for trastuzumab biosimilars. *British Journal of Cancer* (2019). https://doi.org/10.1038/s41416-019-0480-z
2. Chow, SC and Chang, M. (2011). *Handbook of Adaptive Designs in Pharmaceutical and Clinical Development*. Chapman Hall/CRC Press, Taylor and Francis, Boca Raton, FL.
3. Chow, SC and Liu, JP. *Design and analysis of Bioavailability and Bioequivalence Studies*, 3rd ed. USA: Chapman & Hall/CRC Biostatistics Series (2009).

4. Chow, S. C. *Biosimilars: Design and Analysis of Follow-on Biologics. Chapman and Hall/CRC Biostatistics Series.* Taylor and Francis (2013).
5. Chung I., Oh J., Lee S. H., Jang I-J., Lee Y., & Chung J. Y. A post hoc analysis of intra-subject coefficients of variation in pharmacokinetic measures to calculate optimal sample sizes for bioequivalence studies. *Translational and Clinical Pharmacology*, *26*(1), 6-9 (2018). https://doi.org/10.12793/tcp.2018.26.1.6
6. DiMasi, J., Grabowski H. and Hansen, R. Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics* (2016) 47 20-33.
7. DiMasi, J., Grabowski H. and Hansen, R. The Cost of Drug Development. *New England Journal of Medicine,* 372(20):1972, May 2015*.*
8. EMA CPMP Guideline on the Investigation of Bioequivalence
9. EMA Guideline on multiplicity issues in clinical trials
10. FDA Guideline for Industry: Statistical Approaches to Establishing Bioequivalence
11. Fuglsang, A. Sequential Bioequivalence Approaches for Parallel Designs. *American Association of Pharmaceutical Scientists*, Vol. 16, No. 3, May 2014
12. Gould L. (1995). Group sequential extensions of a standard bioequivalence testing procedure. *Journal of pharmacokinetics and biopharmaceutics*, 23(1), 57-85.
13. Haidar S. H., Davit B., Chen M. L., Conner D., Lee L. M., Li Q. H., Lionberger R., Makhlouf F., Patel D., Schuirmann D. J., Yu L. X. Bioequivalence approaches to highly variable drugs and drug products. *Pharm. Res.* 2008, 25:237–241. doi: 10.1007/s11095-007-9434-x
14. Herper, Matthew. The Cost Of Developing Drugs Is Insane. That Paper That Says Otherwise Is Insanely Bad. *Forbes Magazine*, 16Oct2017.
15. Ibrahim, J., Chen, M., Lakshminarayanan, M, Liu, G. and Heysed J. Bayesian probability of success for clinical trials using historical data. *Stat Med*. 2015 January 30; 34(2): 249-264.
16. Jones, B. and Kenward, M. *Design and Analysis of Cross-Over Trials*, 2nd Edition, Chapman & Hall/CRC, Boca Raton, FL, 2003.
17. Labes, D., Schutz, H. and Lang, B. Package 'PowerTOST'. Version 1.4-8 (2019-08-29)
18. Mordenti, J., Pharmacokinetic scale-up: Accurate prediction of human pharmacokinetic profiles from animal data. *J. Pharm.* Sci., 74: 1097-1099, 1985. doi:10.1002/jps.2600741017
19. O'Brien, P. and Fleming, T. A multiple testing procedure for Clinical Trials. *Biometrics*, 1979 Sep;35(3):549-56. doi:10.2307/2530245.
20. Patterson, S. and Jones, B. *Bioequivalence and Statistics in Clinical Pharmacology*, 2nd Ed. Chapman & Hall/CRC, Boca Raton (2017).
21. Pocock, S. Group sequential methods in the design and analysis of clinical trials. *Biometrika*, Volume 64, (1977).
22. Pong A. and Chow S. *Handbook of Adaptive Designs in Pharmaceutical and Clinical Development*. CRC Press (2010).
23. Potvin, D., DiLiberti C., Hauck W., Parr, A., Schuirmann D. and Smith R. Sequential design approaches for bioequivalence studies with crossover designs. *Pharmaceutical Statistics*, 7, 245-262 (2008).
24. Rufibach, K., Burger, H. and Abt, M. (2016). Bayesian predictive power: choice of prior and some recommendations for its use as probability of success in drug development. *Pharmaceutical statistics*, 15(5), 438-446.
25. Schwartz TA and Denne JS. Common threads between sample size recalculation and group sequential procedures. *Pharmaceutical Statistics,* 2, 263–271. DOI: 10.1002/pst.068
26. Spiegelhalter, D. J., Abram, K., Myles, J. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. New York: John Wiley and Sons.
27. Sullivan, Thomas. A Tough Road: Cost To Develop One New Drug Is $2.6 Billion; Approval Rate for Drugs Entering Clinical Development is Less Than 12%. *Policy & Medicine*, 21Mar2019.

28. Tothfalusi, Laszlo and Endrenyi, Lazlo. Sample Sizes for Designing Bioequivalence Studies for Highly Variable Drugs. *Journal of Pharmacy and Pharmaceutical Sciences*. 2012: 15(1) 73-84.
29. Tothfalusi, L. and Endrenyi, L. Algorithms for evaluating reference scaled average bioequivalence: power, bias, and consumer risk. *Statistics in Medicine*. 2017; 36: 4378-4390.
30. Wang, Y., et. al. (2013). Evaluating and utilizing probability of study success in clinical development. *Clinical Trials*, 10(3), 407-13.
31. https://www.fda.gov/drugs/questions-answers/generic-drugs-questions-answers
32. https://www.drugpatentwatch.com/blog/opportunities-for-generic-drug-development/
33. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=320.1