Universidade de Lisboa

Faculdade de Farmácia

Narrow therapeutic range drug products. Challenges in the assessment of Bioequivalence. The role of Modelling and Simulations and expert systems.

Rita Bento Guerreiro

Mestrado Integrado em Ciências Farmacêuticas

2020

Universidade de Lisboa

Faculdade de Farmácia

Narrow therapeutic range drug products. Challenges in the assessment of Bioequivalence. The role of Modelling and Simulations and expert systems.

Rita Bento Guerreiro

Monografia do Mestrado Integrado em Ciências Farmacêuticas apresentada à Universidade de Lisboa através da Faculdade de Farmácia

Orientador: Professor Auxiliar, Luís Filipe Baptista Pleno Gouveia

Coorientador: Professor Auxiliar, Paulo Jorge Pereira Alves Paixão

2020

Resumo

A disponibilização de medicamentos bioequivalentes veio ampliar o leque de escolhas do doente na hora de comprar os medicamentos de que necessita. No entanto, veio também suscitar questões inerentes aos critérios regulamentares para aprovação destes medicamentos. Um caso particular dentro dos medicamentos genéricos são os medicamentos de margem terapêutica estreita, uma vez que estes apresentam riscos acrescidos pela proximidade entre efeitos tóxicos extremos e a eventual falta de efeito terapêutico desejado. Por esta razão, quer a EMA, quer a FDA apresentaram critérios de bioequivalência especialmente pensadas para este tipo de fármacos. Pela relevância deste tema, torna-se imperativo o estudo do impacto da variabilidade no número de indivíduos necessários, por forma a tirar conclusões acerca do impacto desta variabilidade, bem como o risco de conclusão errada dessa mesma bioequivalência em cada um dos contextos regulamentares.

O objetivo deste trabalho consiste na realização de uma pesquisa bibliográfica, seguida da criação de um modelo que aproxime as metodologias já utilizadas por ambas as autoridades reguladoras para concluir acerca da bioquivalência. Para tal, foi necessário o estudo da relação entre a variabilidade intra-individual e o número de indivíduos necessários para demonstrar bioequivalência de acordo com o novo modelo. Para além disso, foi também avaliado o risco de concluir erradamente essa mesma bioequivalência, neste mesmo modelo. Para tal foi utilizado o software R e o pacote PowerTOST como forma de testar a performance de cada do modelo criado.

Palavras-chave: Bioequivalência, Biodisponibilidade, Medicamentos de Margem Terapêutica Estreita, Medicamento Genérico, Critérios de Decisão.

Abstract

The availability of bioequivalent medicines has expanded the patient's range of choices regarding buying the medicines they need. However, it also raised questions regarding the regulatory criteria for the approval of these drugs. A particular case within generic drugs is drugs with a narrow therapeutic index since they present increased risks due to the proximity between extreme toxic effects and the possible lack of desired therapeutic effect. For this reason, both the EMA and the FDA presented bioequivalence criteria specially designed for this type of drug. Due to this theme's relevance, it is imperative to study the impact of variability on the number of individuals required to conclude about the impact of this variability and the risk of the wrong conclusion of the same bioequivalence in each of the regulatory contexts.

This work aims to carry out a bibliographic search, followed by the creation of a model that approximates the methodologies already used by both regulatory authorities to conclude about bioequivalence. Thus, it was necessary to study the relationship between intraindividual variability and the number of individuals needed to demonstrate bioequivalence according to the new model. Also, the risk of wrongly concluding this same bioequivalence was assessed in this same model. In this case, the R software and the PowerTOST package were used to test the performance of each model.

Keywords: Bioequivalence, Bioavailability, Narrow Therapeutic Index Drugs (NTID), Generic Drugs, Decision Criterion.

Acknowledgment

To mom and dad for all their support and for believing in me more than I do. They were tireless, always—Thank you for always cheering for me and for being my favorite fans.

To brother and sister-in-law for the example of dedication and strength in the low moments. For always welcoming me and distracting me in the best and worst moments.

To the grandparents, because they are my biggest example, I am who I am, and I have what I have because of them. They are a great example of work and glory.

To Laura, because without her, it wouldn't have been the experience it was. For all the support, for all the advice, for being my “mother” in carrying out this monography. In desperation times, she never let me give up and always motivated me to stay healthier.

To Bibi, Nardo and Joãozinho for being there, always, at all times. You know you are those friends, always, in my heart.

To Sara and Samuel, for being tireless sponsors who, even having their own jobs, never forgot me and gave me all the support I needed.

To Andreia, who always tried to push me to evil plans but always gave great advice.

To Isi, who gave me “shelter” and comfort to spend these last months doing this hard work.

To Maria and Afonso, for being my children, for whom I have a strong affection. They will always be mine, and I will always be here for them.

To Professor Luís Gouveia and Professor Paulo Paixão for all the guidance given and for having challenged me to develop this fantastic work. You are the best!

To you all, because without you his whole project would not have been possible,
Many thanks!

Index

Table Index

Figure Index

Equation Index

Abbreviation Index

AIM - Marketing Authorization

ANDA - Abbreviated New Drugs Application

AUC - Area under the curve

BCS - Biopharmaceutical Classification System

BD - Bioavailability

BE - Bioequivalence

CI - Confidence Interval

C_{max} – Maximum drug concentration in blood

$C_{max, ss}$ – Maximum drug concentration in blood, at steady state

CV - Coefficient of Variation

DCI - International Common Name

EMA - European Medicines Agency

FDA - US Food and Drug Administration

HVD – Highly Variable Drugs

INFARMED, I.P. - National Authority for Medicines and Health Products, I.P.

NTIDs - Narrow therapeutic index drugs

WHO - World Health Organization

PK - Pharmacokinetics

PD - Pharmacodynamics

RP - Reference product

SNS - National Health System

t_{max} - Time to maximum plasma drug concentration in blood

TOST - Two One-sided Testing Procedure

TP - Test product

2x2x2 – single dose, 2 treatments, 2 sequences, 2 periods, cross-over study

2x3x3 – single dose, 2 treatments, 3 sequences, 3 periods, partial replicate, cross-over study

2x2x4 – single dose, 2 treatments, 2 sequences, 4 periods, full replicate, cross-over study

Introduction

The advances made in the area of health, in terms of innovative technology, are constant and rampant, but this has not prevented the continuous and permanent emergence of questions and problems for which, however, there are still no satisfactory or adequate answers. And the only reliable way to obtain these answers is through continuous studies and scientific validtions.

The pharmaceutical industry and the entities responsible for the approval of medicines have faced two themes that, due to their extreme complexity, relevance, and close interconnection, have been the subject of extensive research and debate: the therapeutic variety and the availability of bioequivalent medicines.

These realities introduced a more excellent supply of medicines on the market and, consequently, allowed the patient to have a broader range of options, at lower prices, at the time of purchase. However, they simultaneously came to foster other questions and doubts among the various entities related to the medicines, namely by those who create, produce, approve and commercialize them, as they are the ones who have to ensure the safety, efficacy, and quality of the medicines produced and marketed.

To understand how the issues listed above are connected, a brief review of the basic concepts associated with therapeutic variety and bioequivalence follows, especially since they are both indispensable for understanding the study carried out here.

1. Bioequivalence and Bioavailability

Two drug products, with the same active substance, therefore chemically equivalent, are considered bioequivalent if they are pharmaceutical equivalents or pharmaceutical alternatives and if their bioavailability (in rate and extent) after administration of the same molar dose are within the predefined acceptable limits, to guarantee similarity in terms of safety and effectiveness. (1)

Bioavailability (BA) is a measure of the rate and extent of a drug that becomes available at the site of action after administration in a therapeutic dose, based on the plasma concentrations obtained after drug administration.(2)

a) Historic Perspective

In 1938, the US Food and Drug Administration (FDA) defined the concept of "new medicine", which led to new safety documentation rules for its manufacturers, with the agency.(3) Subsequently, the notion of "medicine not new, "which is based on the general safety recognition of a branded medicine already on the market. In the 1950s and 1960s, there was an abrupt growth in new drugs and "non-new" drugs that prompted the passage of laws that prohibit the replacement of a prescribed branded drug. At the same time, the Thalidomide disaster occurred, which, in 1962, led to the approval of an amendment to the Kefauver-Harris Drug Amendments Act that requires the manufacturer to demonstrate standards of effectiveness for the new medicine, in addition to safety standards already imposed.(3) In the 1970s, the rising cost of branded drugs led to the repeal of the previously established law, which aimed at prohibiting the substitution of branded drugs for generic drugs.(3) However, many generic drugs have shown problems of bioavailability for which the FDA Abbreviated New Drugs Application (ANDA) was established - in which the bioequivalence of the generic medicine regarding the innovator drug has to be proven - in the form of exemption from carrying out preclinical tests and clinical studies on safety and efficacy, previously determined for the corresponding innovator drugs (3). Therefore, it was necessary to establish the concept of bioequivalence, as well as the development of the Biopharmaceutical Classification System (BCS) and its applications after the regulatory guidelines (4).

In 1984, the Hatch-Waxman Act was created, the primary purpose of which was to establish a system for regulating generic drugs, with the registration of all approved drugs after 1962.(5) After multiple discussions and input from the Universities and Pharmaceutical Industries, the FDA has established that, for demonstration of bioequivalence, generic drugs must result in pharmacokinetic profiles - assessed using the AUC and C_{max} parameters - that are not significantly different from those obtained for the reference drug, accepting a 20% difference between both.(3) However, as the data should be log transformed prior to the statistical analysis, from this difference value of 20% when they are transported to the average scale, they vary between 20% and 25%: (ln (0.8) = - 0.223 and ln (1.25) = 0.223).

In 1989, the generics scandal emerged and brought new suspicions regarding the safety and efficacy of these drugs, resulting in the issuing of a final guidance to the industry, entitled "Bioavailability and Bioequivalence Studies for Orally Administered Drug Products-General Considerations". This guidance was based on a careful analysis of all the recommendations of the various experts and observations from the public and was subsequently issued by the FDA.(3)

In 1992, the FDA determined that bioequivalence testing is the standard approach to conclude that a generic drug is a therapeutic equivalent, compared to the reference drug. At the same time, the "two one-sided testing procedure" (TOST), a single test procedure, appeared, whose purpose is to compare a test formulation (T) with a reference (R). The typical study design is a crossover, random design (two treatments, in two periods, with two sequences - TR vs. RT) applied to regular healthy volunteers, which can be adapted to other designs, as long as the necessary adjustments are made to it. The TOST is equivalent to the calculation of the confidence intervals of 90% of the ratios of the parameters C_{max} and AUC, between the Test and the Reference, with acceptance limits between 0.80 and 1.25, applicable internationally, despite some adaptations by the regulatory entities, and whose simulations and historical practice have confirmed their usefulness in protecting public health.(6) This ensures that the chance of concluding bioequivalence for two formulations differing more than 20% is less than 5% (Type-I error).

b) Bioequivalence Study Design (1)

All bioequivalence studies must be planned in order to best differentiate the effects of the formulation from other effects.

The standard design that applies in this type of study consists of comparing two formulations in a single dose, through the random crossing of two periods and two crossover sequences – a 2x2x2 design. In this type of design, the intra-subject variability can only be roughly estimated, based on the residual variability. Other designs may consider the repetition of administration of the reference product – a partial replicate study, like the 2x3x3 design – or both the test and the reference products – a full replicate study, like the 2x2x4 design. In these cases, the intra-subject variability may be directly estimated.

Besides, the treatment periods must be separated in time, to ensure that the drug concentration administered in the first test period, is below the lower limit of quantification when administered at the beginning of the second period, ensuring total independence between the two. This objective is usually guaranteed after a period of, at least, 5 elimination half-lives.
The number of subjects included in the clinical study should, preferably, be sufficient for allowing the conclusion of bioequivalence considering an expectable acceptable difference (for example, a 5-10% difference) within the expectable variability of the drug product, and its number should be determined based on a power analysis.

Parameters to be investigated

In the European regulatory environment, the determination (assessment of bioequivalence after administration of a single dose should be made by comparing the parameters AUC_{0-t} and C_{max} obtained in the two formulations. In both cases, two formulations are considered to be bioequivalent if the statistical comparison of the Test-to-Reference parameter ratio results in a 90% confidence interval contained within the 80.00-125.00% acceptance interval.

The evaluation of the comparison between the two formulations t_{max} is not necessary unless the drug product is a fast-release formulation that claims one of the following criteria:
- Be clinically relevant;
- Start of relevant action;
- Have associated adverse effects.

In this case, the comparison between the two products should not show any apparent difference in the median of the observed t_{max} values and its variability between the Test and Reference formulations.

Data Analysis

Data analysis must be performed using ANOVA tests. The General Linear Model (GLM) used in ANOVA must include all effects considered relevant, and that may influence the conclusions of the study. In typical trials (crossover, randomized, two-period, single dose), these effects are typically the formulations, the period of administration, the sequence, and the volunteers nested in the sequence effect. Also, there must be a rigorous and easy-to-interpret

record of all data relating to concentration and pharmacokinetic parameters - by formulation and by individual - as well as data on the study randomization scheme. All calculated PK parameters must be accompanied by descriptive statistics.

2. Reference Drugs Versus Generic Drugs

A reference medicine is, generally, the innovator drug associated with high investments on the part of the Pharmaceutical Industry.

On the other hand, the generic drug can prove to be a therapeutic equivalent based on criteria of therapeutic equivalence and bioequivalence. It must present the same active substance in the same pharmaceutical form, dose, and route of administration, in addition to the same potency and concentration, concerning the reference drug. However, it may contain other excipients, since these are considered to be inert substances, without biological activity, as long as they have already been used in other approved drugs, whose efficacy and safety have not been affected.(7)

According to both FDA and EMA models, a drug is considered generic if it exhibits an equivalent concentration-time profile in the blood. In other words, bioequivalence can be established based on C_{max}, T_{max}, and AUC. The acceptance criteria consist of 90% confidence intervals (CIs) and bioequivalence limits, established for the ratio between the test and the reference averages for AUC and C_{max}, which are in the range of 80.00-125.00 %.(7)

3. Narrow Therapeutic Index Drugs

According to the FDA, narrow therapeutic index drugs (NTIDs) are those in which small variations in dose or blood concentration can lead to dangerous therapeutic failures or adverse reactions that can culminate in hospitalization, disability, or even death. Some examples of this type of medication, according to the FDA, are warfarin, levothyroxine, carbamazepine, digoxin, lithium carbonate, phenytoin, and theophylline. The EMA, on the other hand, does not present a clear definition of NTI, stating that it must be evaluated on a case-by-case basis and based on clinical criteria, with some subjectivity inherent in the existence of multiple regulatory agencies.(1) In Portugal, and for reasons of prescription, cyclosporins,

levothyroxine sodium, and tacrolimus are considered NTID. For these drugs, substitution on the prescription is not authorized.(8)

Although the criteria for considering a drug as NTID are subjective, these drugs generally have a narrow therapeutic index and can easily have no therapeutic effect or have toxic concentrations, as well as low intra-individual variability. For this reason, they are subject to therapeutic monitoring by PK or PD parameters in order to ensure safety and efficiency during their use.(9)

4. Bioequivalence Decision Criteria for NTIDs

Within a population, variability can be observed among different subjects, which is called interindividual variability. In addition, the same person may present different responses to the same drug, at the same dose, during the course of the therapy, which is called intraindividual variability. Both inter- and intraindividual variabilities are caused by a combination of factors that affect the drug's PK and PD.(7)

To determine bioequivalence, both variabilities must be considered, with inter-individual variability being a relevant criterion for starting therapy with a given drug, while intraindividual variability is a relevant criterion for changing the type of formulation used.(10)

In the case of NTIDs, they can have significant interindividual variability, responsible for the PK variability of the drug, but often low intraindividual variability. For this reason, the typically acceptable difference of 20% between the formulations under evaluation is reduced to a difference value considered acceptable of only 10%. This difference is typically assessed for the AUC parameter, but also for the C_{max} parameter.(7)

FDA's Model versus EMA's Model

In the model implemented by the FDA, the limits of bioequivalence are changed due to the intraindividual variability of the Reference product - average bioequivalence on a reference

scale ("ABE on a reference scale"). In this way, two different situations arise that encourage the development of formulations with low variability:

→ If the variability (ISCV) associated with the Reference drug is ≤ 10%, the BE limits will be dimensioned for reference and will be narrower than 90-111.11%

→ If the variability (ISCV) associated with the Reference drug is > 10%, the BE limits will have a reference scale and will be greater than 90-111.11%, but will be limited between 80-125%.(11)

The proposal for NTID potency specifications for generic versions aims to ensure that the switch between branded and generic drug products, as well as between generic drug products, provides comparable PK profiles and consistency of the dose administered over the expiration date.(11)

On the other hand, the EMA uses a simpler strategy. For NTIDs, the tighter acceptance range is used when compared to other drugs, which is 90.00-111.11%.(1)

5. Perspectives of the approaches used by the EMA and the FDA

The topics covered so far lead us to question the advantages and disadvantages of each of the protocols presented.

Although there are considerable differences, what will be the validity of each model? What is the reason for FDA-approved NTI drugs that leads to failing approval through the EMA protocol? Is it questionable whether the protocols are too demanding concerning the number of individuals? Could it be that, if we reduce the number of volunteers in the study, we put at risk the approval of NTID drugs with differences larger than 10% between them? What are the errors that can result? In economic terms, is it possible to incur the costs inherent in the number of volunteers needed in the study and keep them loyal until the end? Can an alternative criteria, that optimize the pros/cons relationship of both approaches, be proposed?

These are pertinent questions, frequently asked by the expert in the field, and evaluated during the monography.

Objectives

Bioequivalence decisions are crucial for the approval of generic drugs, especially the approval of drugs considered to have a narrow therapeutic margin. However, the model applied by regulatory authorities, such as the FDA and EMA, differ from each other and, as such, have strengths and weaknesses when compared.

One of the parameters that diverges in the models applied by the two entities is the number of volunteers needed to demonstrate the bioequivalence of generic drugs with a narrow therapeutic margin, making it important to study the impact of their intra-individual variability.

After conducting this study, it is imperative to create a new model for approval of generic drugs with a narrow therapeutic margin, in an attempt to harmonize both the EMA and FDA models.

1. Specific objectives

To carry out this study, it was necessary to acquire knowledge about RStudio software, as well as the PowerTost package, in order to:

- Ascertain the bioequivalence criteria between generic drugs and innovative drugs;
- Understand the margins applied to drugs with a narrow therapeutic margin;
- Identify the drugs that are currently considered to have a narrow therapeutic margin;
- Understand the risks and factors that can lead to wrong decisions in the BE decision;
- Find out the number of volunteers needed, taking into account the potency of a medication, for bioequivalence studies;
- Understand the impact of the number of volunteers at a socio-economic and ethical level for the approval of NTIDs;
- Identify the risks associated with the patient related to the models applied by the FDA and EMA;
- Conclude about the immediate and long-term consequences for Pharmacoeconomics, the NHS, the Portuguese State and at a practical level.
- Determine the number of volunteers needed, taking into account the potency of a drug, for bioequivalence studies, according to the new model being created.

Materials and Methods

The elaboration of this project was based on the analysis and interpretation of data and information collected from several original research and review of scientific articles, as well as information available on a considerable number of scientific sites on the internet.

The sources to obtain electronic bibliography were the platforms: *PubMed* (https://www.ncbi.nlm.nih.gov/pubmed); WHO (https://www.who.int/); EMA (https://www.ema.europa.eu/en); FDA (https://www.fda.gov/home); and INFARMED, I.P. (https://www.infarmed.pt/). This research was carried out in the period between the June 30th2019 and the October 31th2020.

To perform the comparative study of power with the number of individuals, in both models, RStudio was used, a statistical software whose purpose is to develop and share your work on a large scale. In this software, the PowerTost package, which contains a number of functions to calculate the power and the sample size for several bioequivalence study designs (https://cran.r-project.org/web/packages/PowerTOST/ PowerTOST.pdf), was included.
The results obtained were exported from RStudio to Microsoft Excel files, which allowed further processing and analysis of the obtained data.

The study carried out, in a first phase, evaluated the number of volunteers needed to demonstrate BE and, in a second phase, the risk of mistakenly concluding about BE.

Subsequently, a new model based on the scaled-average-bioequivalence concept was proposed with new acceptance limits. These were considered fixed at 90-111%, for CVintra values below 13,92%. For higher CVintra values, the extent of the widening is defined based upon the within-subject variability seen in the bioequivalence study according to [U, L] = exp [±k.sWR], where U is the upper limit of the acceptance range, L is the lower limit of the acceptance range, k is the regulatory constant set to 0,760 and sWR is the within-subject standard deviation of the log-transformed values of the reference product parameter under evaluation. This widening is finally caped at 80-125% for CVintra values above 30%.

Statistical tests were carried out in order to assess the validity of the BE decision criteria implemented in the protocols of the regulatory authorities and of the proposed model.

The Hypothesis Test is the method that will be used to study the three protocols under study. Thus, a hypothesis is posed - the probability that the protocols under study will erroneously conclude about the BE of two formulations for approval of NTID drugs - assuming that the alpha error (type I) is 5%.

1. Evaluation of the number of volunteers needed to demonstrate BE

In the first part of the work, a study was carried out that consisted on the evaluation of the number of volunteers needed to demonstrate BE of generic drugs, according to the models applied by the EMA, the FDA and the newly proposed one.

EMA's model

To carry out this phase of the study, according to the EMA model, two different situations were studied:

1) Assuming that the two formulations are the same (theta0 = 1);

2) Assuming that there is a 5% difference between the formulations (theta0 = 0.95).

This procedure was repeated for the three designs used by the EMA model, namely the 2x2x2, 2x3x3 and 2x2x4. In all situations, a study power of 80% was considered.

For each of the situations, CV was varied, with jumps of 1% between them, according to table 1:

Table 1 - Variation of design and CV, depending on the respective theta0, in the EMA's model.

Design	Theta 0		CV	N Gaps
2x2x4	0.95	1	6% - 40%	1%
2x3x3	0.95	1	6% - 40%	1%
2x2x2	0.95	1	6% – 40%	1%

To perform this test the following function 1 was used (12):

$$\text{sampleN.TOST}(\text{alpha} = 0.05, \text{targetpower} = 0.8, \text{logscale} = \text{TRUE}, \text{theta0}, \text{theta1} = 0.9, \text{theta2} = 1.11, \text{C}, \text{design} = \text{2x2x2}, \text{method} = \text{exact}, \text{robust} = \text{FALSE}, \text{print} = \text{TRUE}, \text{details} = \text{FALSE}, \text{imax} = 100) \quad (1)$$

Arguments (12)**:**

Alpha - Type I error probability. Per convention mostly set to 0.05.
Targetpower - Power to achieve at least. Must be >0 and <1. Typical values are 0.8 or 0.9.
Logscale - Should the data used on log-transformed (TRUE) or on original scale (FALSE)? Defaults to TRUE.
Theta0 - 'True' or assumed T/R ratio. In case of logscale=TRUE it must be given as ratio, otherwise as difference to 1. Defaults to 0.95 if logscale=TRUE or to 0.05 if logscale=FALSE.
Theta1 - Lower bioequivalence limit. In case of logscale=TRUE it is given as ratio, otherwise as difference to 1. Defaults to 0.8 if logscale=TRUE or to -0.2 if logscale=FALSE.
Theta2 - Upper bioequivalence limit. If not given theta2 will be calculated as 1/theta1 if logscale=TRUE or as -theta1 if logscale=FALSE.
CV - Coefficient of variation as ratio.
design - Character string describing the study design.
method - Method for calculation of the power.
robust - Defaults to FALSE. With that value the usual degrees of freedom will be used. Set to TRUE will use the degrees of freedom according to the 'robust' evaluation. Has only effect for higher-order crossover designs.
print - If TRUE (default) the function prints its results. If FALSE only the data.frame with the results will be returned.
details - If TRUE the design characteristics and the steps during sample size calculations will be shown. Defaults to FALSE.
imax - Maximum number of steps in sample size search. Defaults to 100. Adaption only in rare cases needed.

FDA's model

To carry out the same study, applied to the FDA model, two different situations were studied:

1) Assuming that the two formulations are the same (theta0 = 1);

2) Assuming that there is a 5% difference between the formulations (theta0 = 0.95);

For each of the situations, CV was varied, with jumps of 2.5% among themselves, according to table 2 and, again, assuming a power of 80%:

Table 2 - Variation of the CV according to the respective theta0, in the FDA model.

Design	Theta 0		CV	N Gaps
2x2x4	0.95	1	6% - 40%	1%

To perform this test the following function 2 was used (12):

$$sampleN.NTIDFDA(alpha = 0.05, targetpower = 0.8, \boldsymbol{theta0}, \boldsymbol{theta1} = \boldsymbol{0.8}, \boldsymbol{theta2} = \boldsymbol{1.25}, \boldsymbol{CV}, design = 2x2x4, nsims = 1e+05, nstart, imax = 100, print = TRUE, details = TRUE, setseed = TRUE) \quad (2)$$

Arguments(12)**:**

Alpha - Type I error probability. Per convention mostly set to 0.05.

Targetpower - Power to achieve at least. Must be >0 and <1. Typical values are 0.8 or 0.9.

Theta0 - 'True' or assumed T/R ratio.

Theta1 - Conventional lower ABE limit to be applied in the FDA procedure. Defaults to 0.8 if not given explicitly.

Theta2 - Conventional upper ABE limit to be applied in the FDA procedure. Defaults to 1.25 if not given explicitly.

CV - Intra-subject coefficient(s) of variation as ratio (not percent).

Design - Design of the study to be planned. "2x2x4" is the full replicate with 2 sequences and 4 periods (TRTR|RTRT). "2x2x3" is the full replicate with 2 sequences and 3 periods (TRT|RTR). Defaults to design="2x2x4".

nsims - Number of simulations to be performed to obtain the empirical power. Defaults to 100,000 = 1e+5.

nstart - Set this to a start value for the sample size if a previous run failed. May be missing.

imax - Maximum number of steps in sample size search. Defaults to 100.

print - If TRUE (default) the function prints its results. If FALSE only the resulting dataframe will be returned.

details - If set to TRUE, the default, the steps during sample size search are shown. Moreover the details of the method settings are printed.

setseed - Simulations are dependent on the starting point of the (pseudo) random number generator. To avoid differences in power values for different runs a set.seed(123456) is issued if setseed=TRUE, the default.

New Model

In this phase of the study, an estimate of the sample size required by calculating the power of the BE decision through scaled expanded BE acceptance limits, based on simulations, was studied.

For this purpose, the constants indicated in table 3 were used.

Table 3 - Values of constants used in the simulation sampleN.scABEL

Constant	Value
Theta0	1 and 0,95
CV	6% - 40%
n	1%

In order to obtain results for both regulatory agencies, the Regulator criterion was varied between EMA and FDA, for the following Design:

1) 2x3x3 – TRR | RTR | RRT

2) 2x2x4 - TRTR | RTRT ; TRRT | RTTR ; TTRR | RRTT

To perform this test the following function 3 was used (12):

$$sampleN.scABEL(alpha = 0.05, targetpower = 0.8, \boldsymbol{theta0} = \mathbf{0.9}, \boldsymbol{theta1} = \mathbf{1.11}, \boldsymbol{theta2}, \boldsymbol{CV}, design = c(2x2x3), \boldsymbol{regulator}, nsims = 1e+05, imax = 100, print = TRUE, details = TRUE, setseed = TRUE) \quad (3)$$

Arguments:(12)

Alpha - Type I error probability. Per convention mostly set to 0.05.

Targetpower - Power to achieve at least. Must be >0 and <1. Typical values are 0.8 or 0.9.

Theta0 - 'True' or assumed T/R ratio.

Theta1 - Conventional lower ABE limit to be applied in the mixed procedure if CVsWR <= CVswitch. Also Lower limit for the point estimate constraint. Defaults to 0.8 if not given explicitly.

Theta2 - Conventional upper ABE limit to be applied in the mixed procedure if CVsWR <=

CVswitch. Also upper limit for the point estimate constraint. Defaults to 1.25 if not given explicitly.

CV - Intra-subject coefficient(s) of variation as ratio (not percent).

Design - Design of the study to be planned.

"2x3x3" is the partial replicate design.

"2x2x4" is a full replicate design with 2 sequences and 4 periods.

Regulator - Regulatory settings for the widening of the BE acceptance limits. May be given as character from the choices "EMA", "HC", "FDA" or as an object of class 'regSet' (see reg_const). Defaults to regulator="EMA" if missing. This parameter will be explained further ahead.

nsims - Number of simulations to be performed to obtain the (empirical) power. The default value 100,000 = 1e+5 is usually sufficient. Consider to rise this value if theta0<=0.85 or >=1.25. But see the warning section.

imax - Maximum number of steps in sample size search. Defaults to 100.

print - If TRUE (default) the function prints its results. If FALSE only the result data.frame will be returned.

details - If set to TRUE (default), the steps during sample size search are shown.

setseed - Simulations are dependent on the starting point of the (pseudo) random number generator. To avoid differences in power for different runs a set.seed(123456) is issued if setseed=TRUE, the default.

2. Assessment of the likelihood of wrongly concluding by BE

In the second phase of the work, a study was carried out to assess the probability of wrongly concluding about BE, taking into account the number of volunteers and the variation coefficient used in the models of both International Regulatory Entities. At this stage, for each method it was tested a 10%, 15% and 20% difference between the two formulations allowing to evaluate the type-1 error in each situation.

EMA's model

To carry out this phase of the study, according to the EMA's model, two different situations were studied, according to table 4:

Table 4 - Variation of CV and n, depending on the study design, in the EMA's model. The variable jumps in n corresponds to the interval between each n, in the indicated interval. The CV function varies in 2.5% intervals.

Design	CV	n	N gaps	Theta0
2x3x3	6% - 40%	15 - 495	15	0.9 ; 0.85 ; 0.8
2x2x4	6% - 40%	10 - 500	10	0.9 ; 0.85 ; 0.8

To perform this test the following function 4 was used (12):

$$power.TOST.sim(alpha = 0.05, logscale = TRUE, \boldsymbol{theta1} = \mathbf{0.9}, \boldsymbol{theta2} = \mathbf{1.11}, \boldsymbol{Theta0}, \boldsymbol{CV}, \boldsymbol{n}, design = \text{2x2x2}, robust = FALSE, setseed = TRUE, nsims = 1e+05) \quad (4)$$

Arguments(12)**:**

Alpha - Type I error probability, significance level. By convention mostly set to 0.05.

Logscale - Should the data used on log-transformed or on original scale? TRUE (default) or FALSE.

Theta1 - Lower bioequivalence limit. In case of logscale=TRUE it is given as ratio, otherwise as a difference to 1. Defaults to 0.8 if logscale=TRUE or to -0.2 if logscale=FALSE.

Theta2 - Upper bioequivalence limit. If not given theta2 will be calculated as 1/theta1 if logscale=TRUE or as -theta1 if logscale=FALSE.

Theta0 - 'True' or assumed T/R ratio. Varied between 0.9, 0.85 and 0.8.

CV - Coefficient of variation as ratio. In case of cross-over studies this is the within-subject CV and in case of a parallel-group design the CV of the total variability.

n - Number of subjects under study. Is total number if given as scalar, else number of subjects in the (sequence) groups. In the latter case the length of n vector has to be equal to the number of (sequence) groups.

Design - Character string describing the study design.

setseed - Simulations are dependent on the starting point of the (pseudo) random number

generator. To avoid differences in power for different runs a set.seed(1234567) is issued if setseed=TRUE, the default. Set this argument to FALSE to view the variation in power between different runs.

nsims - Number of studies to simulate. Defaults to 100,000 = 1E5.

FDA's model

In the second phase of the study, for the FDA model, the 2x2x4 design was applied and the remaining uncertainties were varied, according to table 5.

Table 5 - Variation of CV and n, according to the study design, in the FDA model. The variable jumps in n corresponds to the interval between each n, in the indicated interval.

Design	CV	n	N gaps
2x2x4	6% - 40%	10 - 80	2

To perform this test the following function 5 was used (12):

$$power.NTIDFDA(alpha = 0.05, \boldsymbol{theta1} = \mathbf{0.8}, \boldsymbol{theta2} = \mathbf{1.25}, \boldsymbol{Theta0}, \boldsymbol{CV}, \boldsymbol{n}, design = \text{2x2x4}, nsims = 1e+05, details = FALSE, setseed = TRUE) \quad (5)$$

Arguments(12)**:**

Alpha - Type I error probability, significance level. Conventionally mostly set to 0.05.

Theta1 - Conventional lower ABE limit to be applied in the FDA procedure. Defaults to 0.8 if not given explicitly.

Theta2 - Conventional upper ABE limit to be applied in the FDA procedure. Defaults to 1.25 if not given explicitly.

Theta0 - 'True' or assumed T/R ratio. Varied between 0.9, 0.85 and 0.8.

CV - Intra-subject coefficient(s) of variation as ratio (not percent).

n - Number of subjects under study.

Design - Design of the study to be planned.

nsims - Number of simulations to be performed to obtain the empirical power. Defaults to 100,000 = 1e+5.

details - If set to TRUE the computational time is shown as well as the components for the BE decision.

setseed - Simulations are dependent on the starting point of the (pseudo) random number generator. To avoid differences in power for different runs a set.seed(123456) is issued if setseed=TRUE, the default.

New Model

For the second phase of the study, an estimate of the sample size required by calculating the power of the BE decision through scaled widened BE acceptance limits, based on simulations, was studied.

For this purpose, the constants indicated in table 6 were used.

Table 6 - Values of constants used in the simulation power.scABEL

Constant	**Value**	
CV	6% - 40%	
Number of jumps	3	2
n	15 - 102	10 – 80
Design base	2x3x3	2x2x4

In order to obtain results for both regulatory entities, the Regulator criterion was varied between EMA and FDA, for the following design:

1) 2x3x3 – TRR | RTR | RRT

2) 2x2x4 - TRTR | RTRT ; TRRT | RTTR ; TTRR | RRTT

To perform this test the following function 6 was used (12):

$$power.scABEL(alpha = 0.05, \boldsymbol{theta1} = \mathbf{0.9}, \boldsymbol{theta2} = \mathbf{1.11}, \boldsymbol{theta0}, \boldsymbol{CV}, \boldsymbol{n}, design = c(2x3x3, 2x2x4, 2x2x3), \boldsymbol{regulator}, nsims, details = FALSE, setseed = TRUE) \quad (6)$$

Arguments(12)**:**

Alpha - Type I error probability, significance level. Conventionally mostly set to 0.05.

Theta1 - Conventional lower ABE limit to be applied in the mixed procedure if CVsWR <= CVswitch. Also lower limit for the point estimate constraint. Defaults to 0.8 if not given explicitly.

Theta2 - Conventional upper ABE limit to be applied in the mixed procedure if CVsWR <= CVswitch. Also upper limit for the point estimate constraint. Defaults to 1.25 if not given explicitly.

Theta0 - 'True' or assumed T/R ratio. Varied between 0.9, 0.85 and 0.8.

CV - Intra-subject coefficient(s) of variation as ratio (not percent).

n - Number of subjects under study.

Design - Design of the study.

"2x3x3" is the partial replicate design.

"2x2x4" is a full replicate design with 2 sequences and 4 periods.

Regulator - Regulatory settings for the widening of the BE acceptance limits. May be given as character from the choices "EMA", "HC", "FDA" or as an object of class 'regSet' (see reg_const). Defaults to regulator="EMA" if missing. This argument may be given also in lower case if given as character. The former regulator="ANVISA" is no longer allowed. Since 2016 the ANVISA recommends the EMA' regulatory settings. This parameter will be explained further ahead.

nsims - Number of simulations to be performed to obtain the empirical power. Defaults to 100,000 = 1e+05. If not given and theta0 equals one of the expanded limits (i.e., simulating empirical alpha), defaults to 1e+06.

details - If set to TRUE the computational time is shown as well as the components for the BE decision.

setseed - Simulations are dependent on the starting point of the (pseudo) random number generator. To avoid differences in power for different runs a set.seed() is issued if setseed=TRUE, the default.

Regulator

In this case, the formula below was used, with the values represented in table 7:

Table 7 - Values of constants used in the simulation Regulator.

regulator	"user"	-
R_const	0.76	Regulatory constant.
CVswitch	0.1392	CV to switch to the widened acceptance limits.
CVcap	0.3	CV for capping the widening of the acceptance limits.

Pe_constr	"FALSE"	Logical. Shall pe constraint be applied? Defaults to TRUE. In this case defaults to FALSE.

$$reg\_const(\boldsymbol{regulator}, \boldsymbol{r\_const}, \boldsymbol{CVswitch}, \boldsymbol{CVcap}, \boldsymbol{pe\_constr}) \quad (7)$$

Arguments:(12)

Regulator - Name of the regulatory body as a string. Implemented settings are for "EMA", "FDA", and "HC". The former (inofficial) settings for "ANVISA" are covered by the EMA settings. In case of regulator="USER" the other arguments must be given. Otherwise, they may be missing.

Results and Discussion

1. Assessment of the number of volunteers needed to demonstrate BE

Initially, the work scheme was divided by designs, which could be the full replicate design with 2 sequences and 4 periods – 2x2x4 – the only one allowing to determine the test and reference individual intra-subject variabilities, and required by the FDA , the partial replicate design – 2x3x3 – that allows to determine the reference product intra-subject variability, and should be sufficient for the newly proposed criteria, or a non-replicate design – 2x2x2, that is sufficient for the current EMA criteria. Subsequently, and for each of the designs, it was assumed formulations are the same - theta0 = 1 - or that they have a 5% difference between them - theta0 = 0.95, as shown in figure 1.

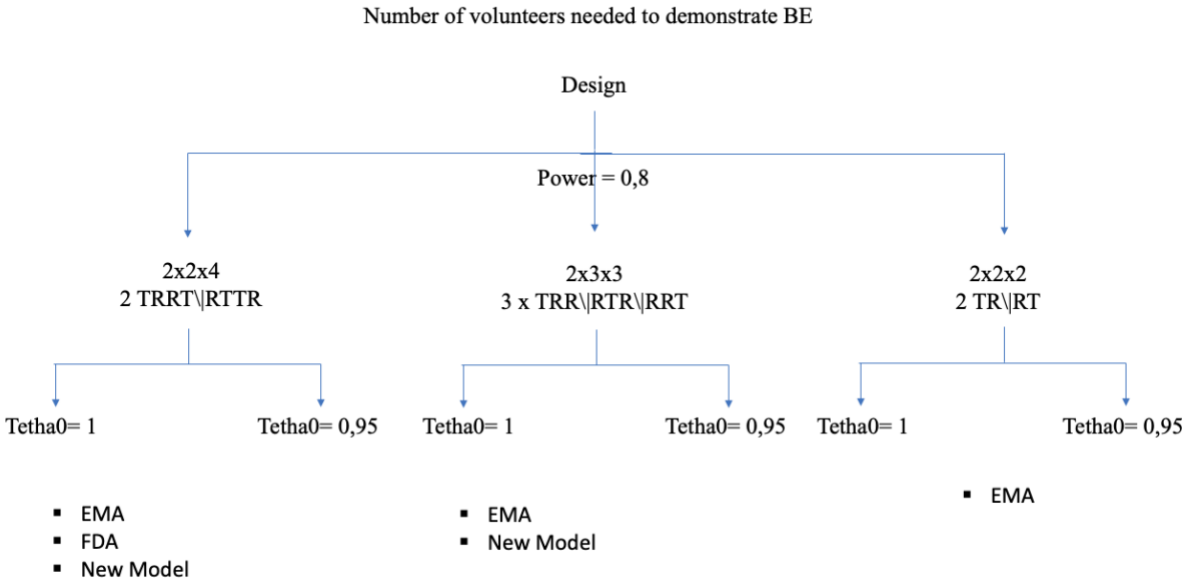

Figure 1 - Scheme that translates the way the results will be presented in the first phase of the study.

Design 2x2x4

Figures 2, 3 and 4 show the results of the first phase of this study, the study of the number of volunteers needed to demonstrate bioequivalence as a function of the coefficient of variation, for the full replicate design with 2 sequences and 4 periods (2x2x4) and theta0 = 1 designs for the EMA, FDA and model under test, respectively, assuming the formulations are equal to each other.

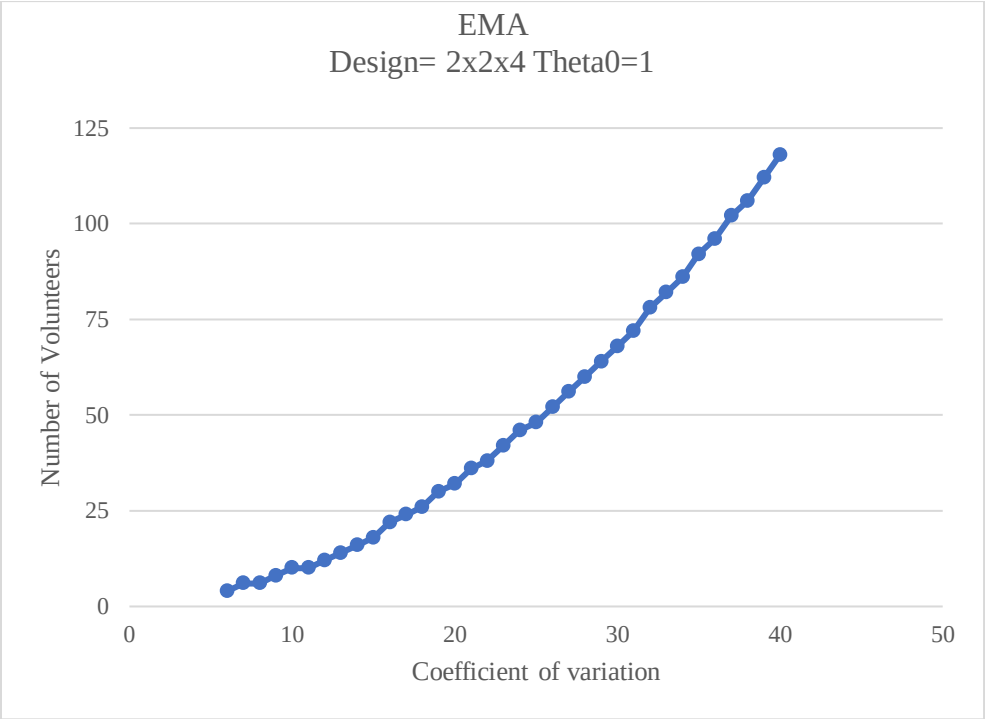

Figure 2 - This figure shows the variation in the number of individuals needed to demonstrate bioequivalence, depending on the coefficient of variation assuming that formulations are equal, according to EMA's model to 2x2x4 design.

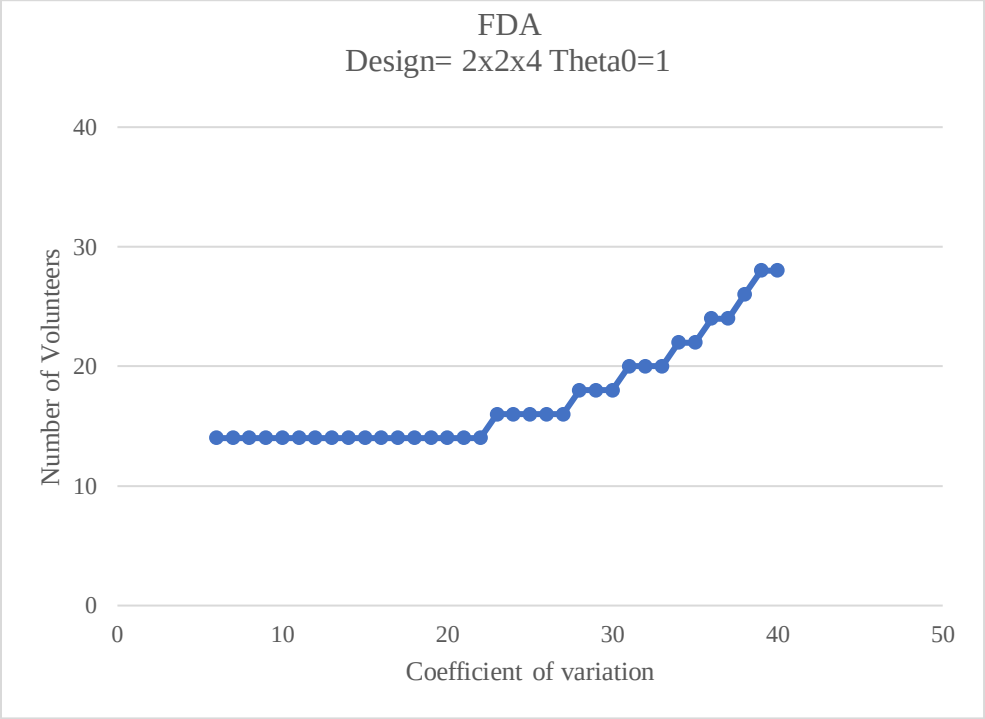

Figure 3 - This figure shows the variation in the number of individuals needed to demonstrate bioequivalence, depending on the coefficient of variation assuming that formulations are equal, according to the FDA model to 2x2x4 design.

Figure 4 - This figure shows the variation in the number of individuals needed to demonstrate bioequivalence, depending on the coefficient of variation, assuming that formulations are equal, according to the New Model to 2x2x4 design.

Through the analysis of figure 2, which represents EMA's models, it is understood that as the coefficient of variation, in other words, the intra-individual variability increases, the number of subjects necessary to demonstrate bioequivalence increases.

Figures 3 and 4 - FDA and New Model, respectively - show that the number of individuals increases on a ladder. Up to values of the coefficient of variation of 30%, 20 individuals are needed, this growth being from 10 to 20 individuals. From values between 30% and 40% of intra-individual variability, the increase is larger and ranges from 20 to 28 individuals necessary to demonstrate bioequivalence.

When compared to each other, it is understood that for the EMA model more volunteers are needed at each point of the coefficient of variation compared to the other two models under study.

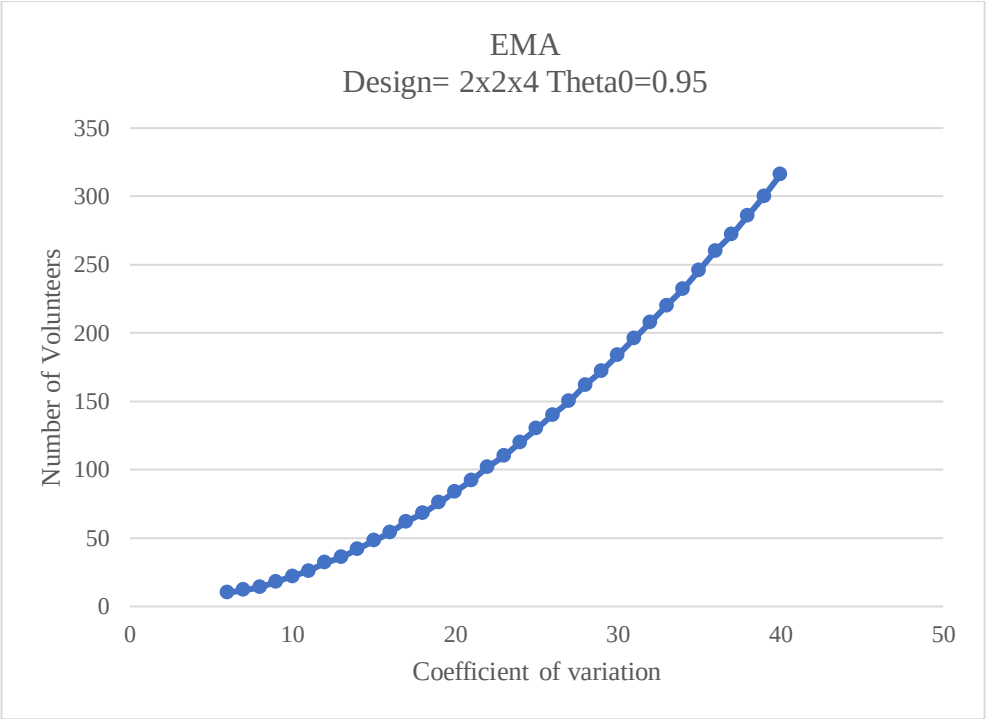

Figure 5 - This figure shows the variation in the number of individuals needed to demonstrate bioequivalence, depending on the coefficientof variation, assuming that the formulations differ by 5%, according to the EMA's model to 2x2x4 design.

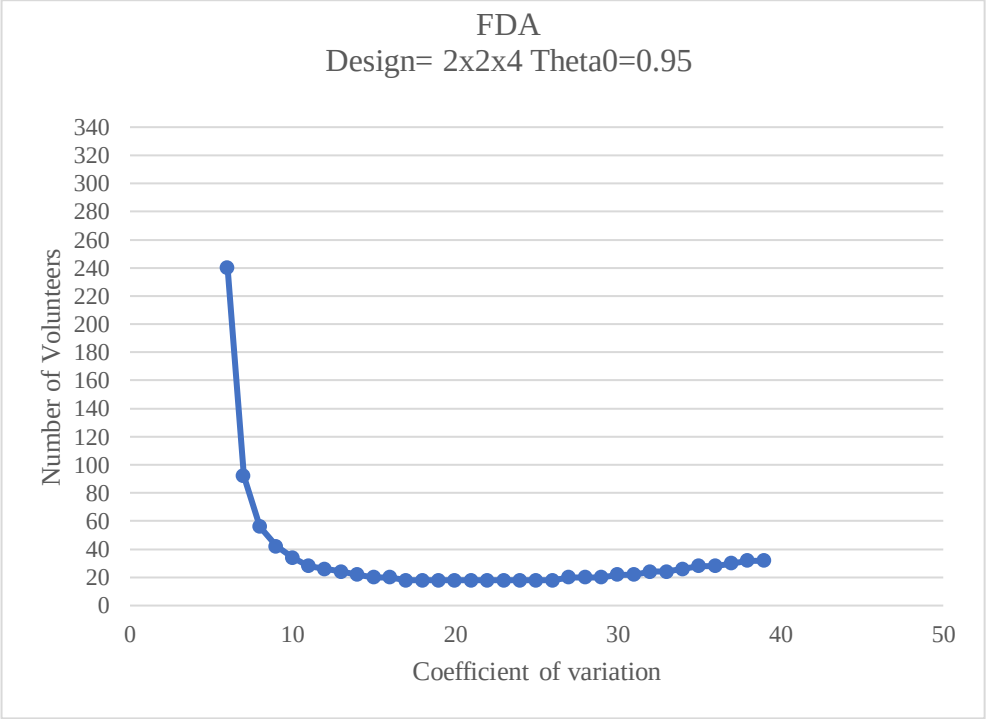

Figure 6 - This figure shows the variation in the number of individuals needed to demonstrate bioequivalence, depending on the coefficient of variation, assuming that the formulations differ by 5%, according to FDA's model to 2x2x4 design.

Figure 7 - This figure shows the variation in the number of individuals needed to demonstrate bioequivalence, depending on the coefficient of variation, assuming that the formulations differ by 5%, according to the New Model to 2x2x4 design.

For the same design, 2x2x4, but assuming a variation of 5% between formulations, figures 5, 6 and 7 were obtained for the EMA, FDA and New Model models, respectively.

Through the analysis of figure 5, it is possible to conclude that there is more than proportional growth in the number of individuals needed to demonstrate bioequivalence for EMA's model. These values are higher for each point of the coefficient of variation than the corresponding graph obtained for equal formulations.

Figure 6, referring to the FDA model, shows a decrease in the number of individuals up to a coefficient of variation of 30%, increasing afterwards for CV larger that 30%.

The New Model results, represented in figure 7, demonstrate that, up to a 15% coefficient of variation, the number of individuals needed in the study to demonstrate bioequivalence increases to values of 40. Then, there is a decrease in this number of individuals to 24, up to

31% of intra-individual variability. These numbers rise again from 32%, reaching values of 35 individuals for a coefficient of variation of 40%.

Design 2x3x3

In the second phase of the study, the EMA and New Model models' results are observed according to the 2x3x3 design, assuming equal formulations and, later, assuming that the formulations differ by 5%.

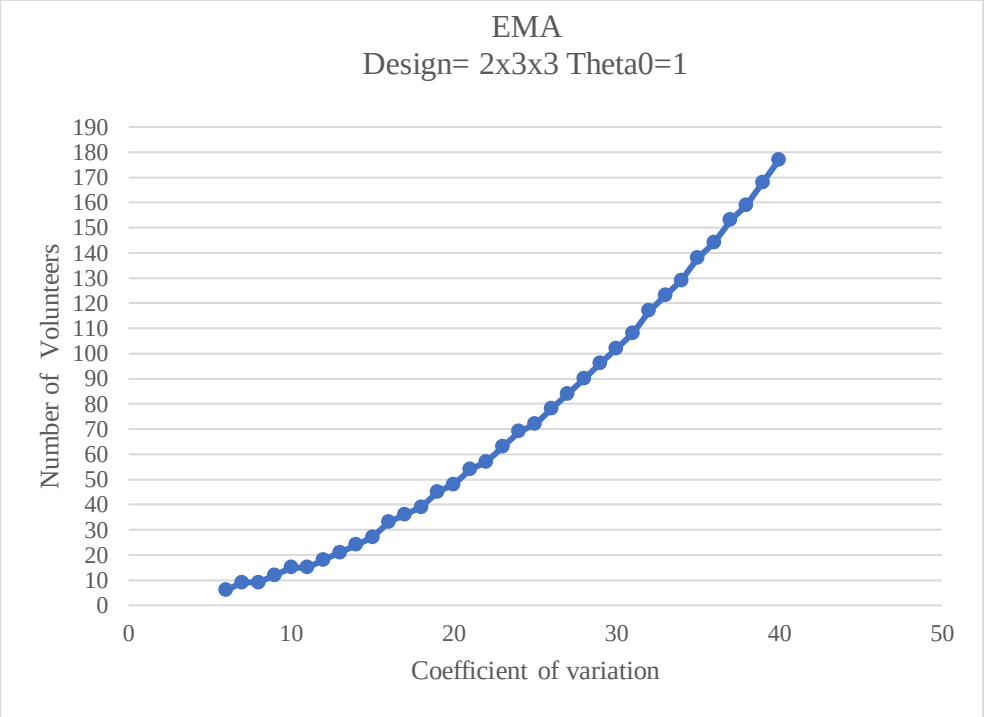

Figure 8 - This figure shows the variation in the number of individuals needed to demonstrate bioequivalence, depending on the coefficient of variation assuming that formulations are equal, according to EMA's model to 2x3x3 design.

Figure 9 - This figure shows the variation in the number of individuals needed to demonstrate bioequivalence, depending on the coefficient of variation, assuming that formulations are equal, according to the New Model to 2x3x3 design.

For the 2x3x3 design, assuming that the formulations are the same, it is observed that in figure 8, the number of individuals required to conclude bioequivalence in the EMA model greatly increases reaching values of 180 individuals for coefficients of variation of 40%.

In figure 9, the New Model results present a value variation, with 42 being the maximum number of individuals required in the study, for an intra-individual variability of 40%.

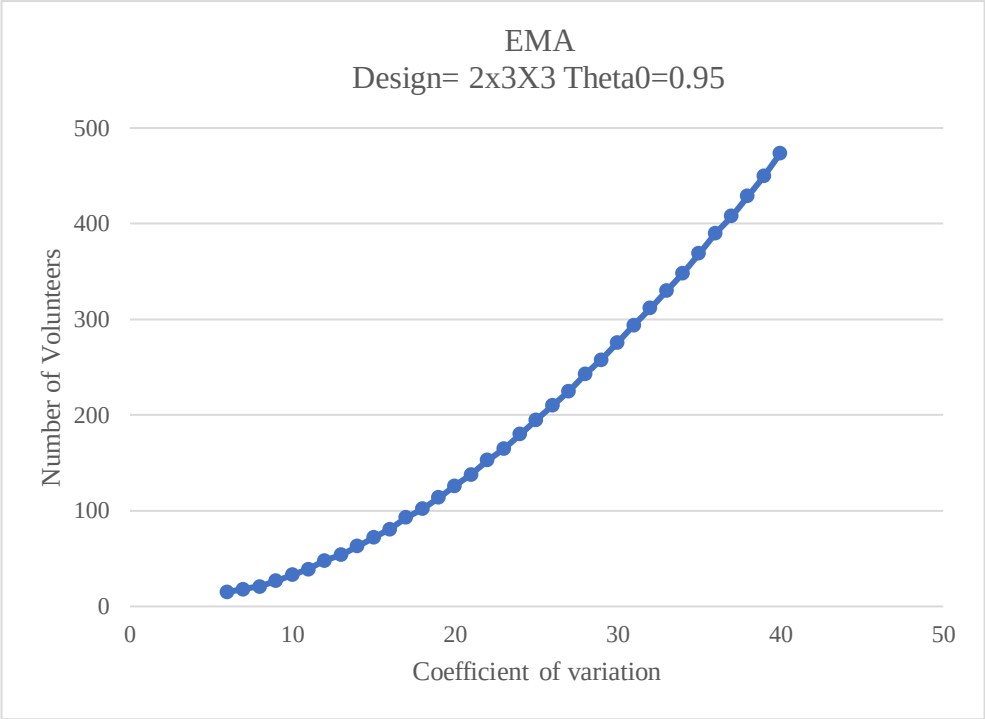

Figure 10 - Variation in the number of individuals needed to demonstrate bioequivalence, depending on the coefficient of variation, assuming that the formulations differ by 5%, according to the EMA model to 2x3x3 design.

Figure 11 - Variation in the number of individuals needed to demonstrate bioequivalence, depending on the coefficient of variation, assuming that the formulations differ by 5%, according to the New Model to 2x3x3 design.

For the same design, 2x3x3, assuming that the formulations vary by 5%, figures 10 and 11 were obtained.

The analysis of figure 10 allows to understand that similar to what was seen in the other designs, the number of individuals required to conclude bioequivalence in the EMA model grows more than proportionally to highly coefficient of variation. This parameter reaches values of 480 individuals for coefficients of variation of 40%.

Figure 11 represents the New Model where the estimated number of individuals needed to demonstrate bioequivalence, up to a coefficient of variation of 15%, grows up to 60. Then, there is a decrease in this number of individuals to 35, up to 30% intra-individual variability. Finally, these numbers rise again from 31%, reaching 50 individuals' values for a coefficient of variation of 40%.
These results are similar to those obtained for the 2x2x4 design, assuming the same difference between formulations.

Design 2x2x2

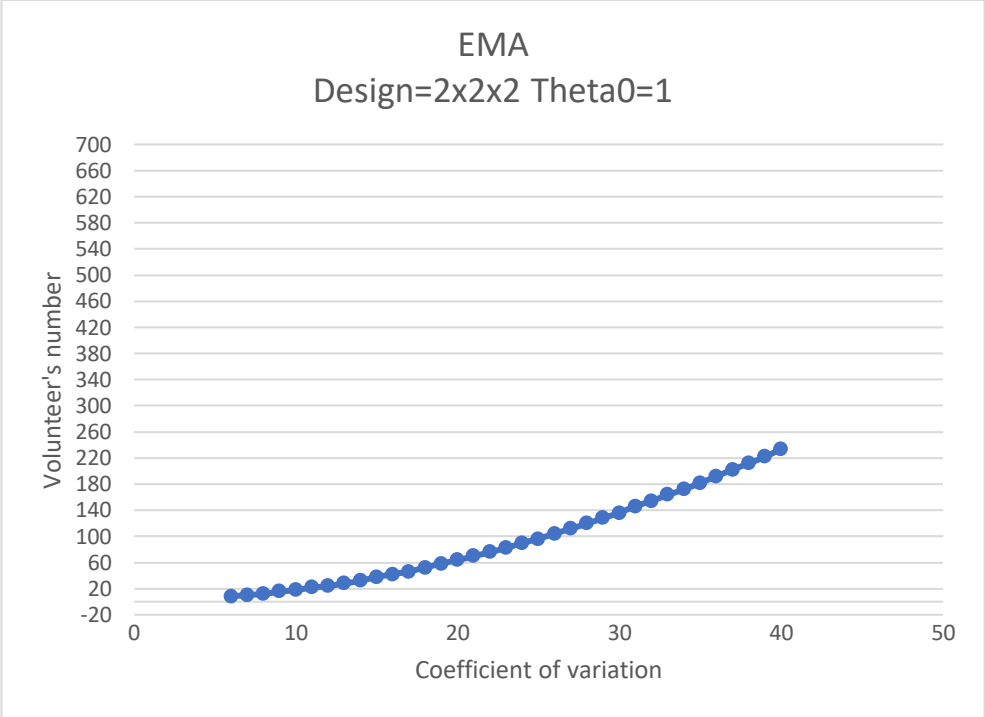

Figure 12 - Variation in the number of individuals needed to demonstrate bioequivalence, depending on the coefficient of variation, assuming that the formulations are equal, according to EMA and 2x2x2 design.

In figure 12, representative of the EMA model, assuming the random crossing of two periods and two crossover sequences design (2x2x2), we observe a value of the number of individuals needed in the study to demonstrate BE of about 240 individuals was observed, assuming that the formulations are the same.

For the same design, assuming a difference of 5% between formulations, we obtained the results presented in figure 13 were obtained.

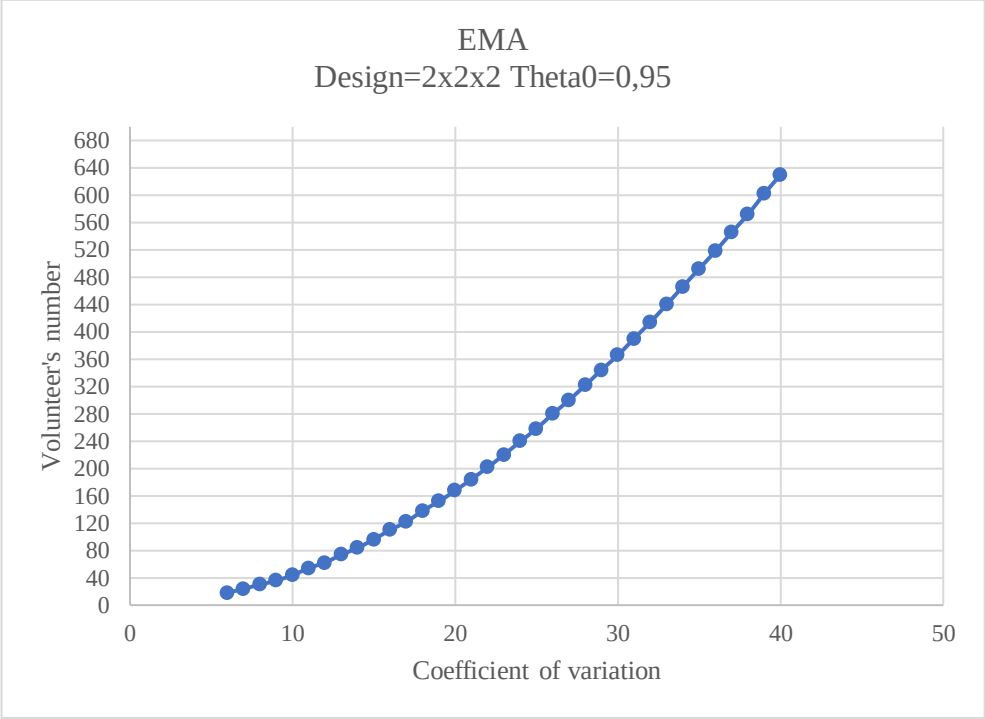

Figure 13 - Variation in the number of individuals needed to demonstrate bioequivalence, depending on the coefficient of variation, assuming that the formulations differ by 5%, according to EMA and 2x2x2 design.

Figure 13 shows a more than proportional growth in the numbers of voluntaries needed in study to demonstrate bioequivalence - when compared to the same model, assuming the same design but equal formulations - reaching values of 640 subjects for a coefficient of variation of 40%.

2. Assessment of the likelihood of wrongly concluding by BE

The purpose of the second part of the study was to evaluate the probability of BE's wrong conclusion. In order to achieve that, a simulation of the probability of accepting a BE

conclusion by considering a real difference of 10%, 15% and 20% between the test and reference drug products was carried out, according to the scheme shown in figure 14.

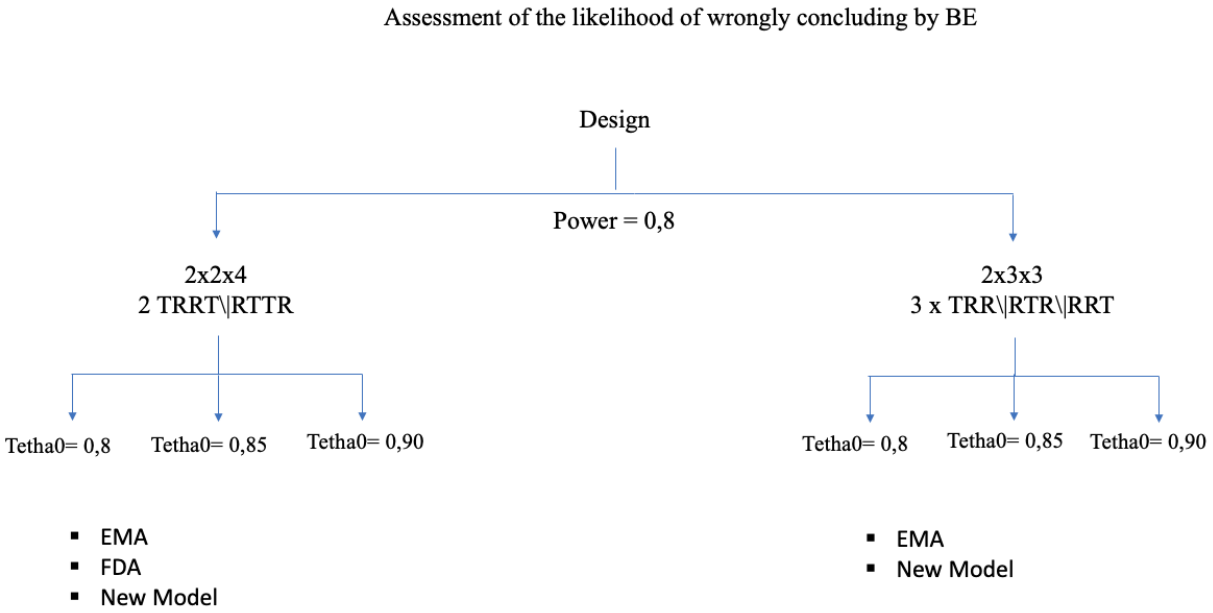

Figure 14 - Scheme that translates the way the results will be presented in the second phase in the study.

The error values in a statistical test are marked as follows:

- Values below 0.05 marked in blue - acceptable error;
- Values between 0.05 and 0.1 marked in green;
- Values above 0.1 marked in red.

The alpha error (type I) was set to an acceptable level of 5%. Thus, when testing a Theta0=0.9, it would be expected that less than 5% of studies with formulations whose differences are larger than 10% would result in a conclusion for bioequivalence. The graphs below represent these conclusions and whether they are acceptable or not, using the symbology mentioned above.

Design 2x2x4, Theta0=0,9

Figures 15, 16 and 17 represent the 2x2x4 design, assuming a 10% difference between the formulations.

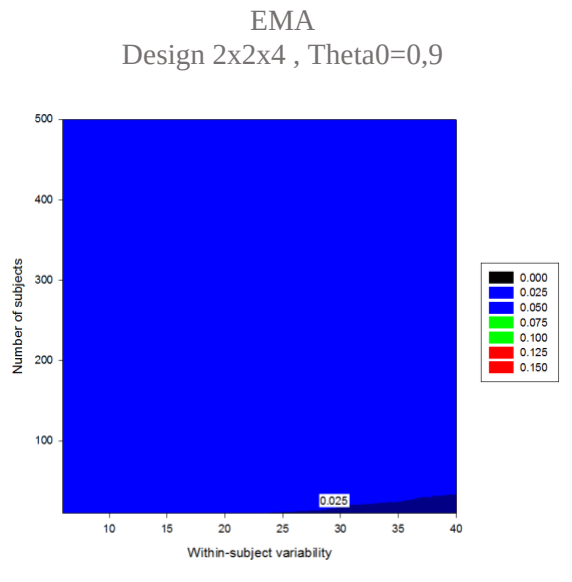

Figure 15 - EMA. Evaluation of the probability of wrongly concluding by BE, assuming Theta0 = 0.9, in the 2x2x4 design.

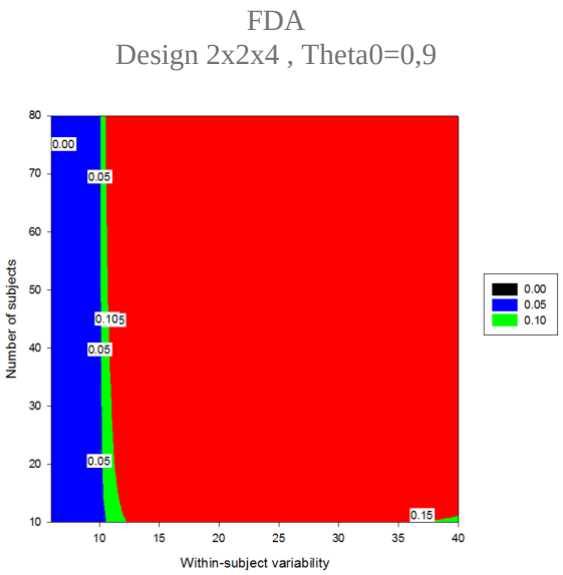

Figure 16 - FDA. Evaluation of the probability of wrongly concluding by BE, assuming Theta0 = 0.9, in the 2x2x4 design.

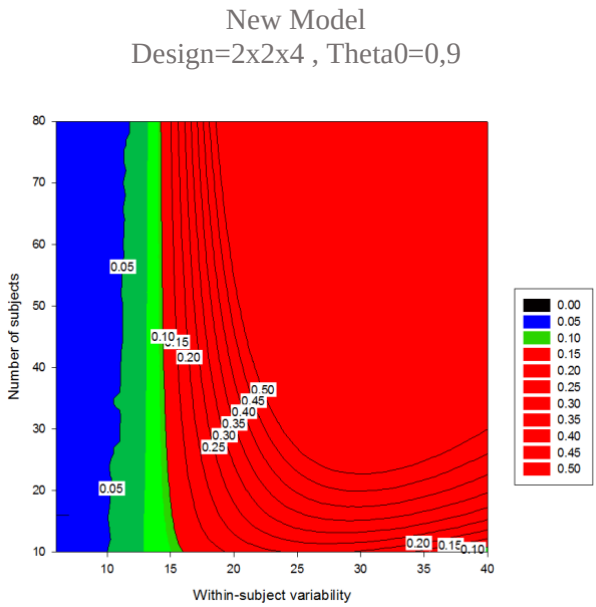

Figure 17 - New Model. Evaluation of the probability of wrongly concluding by BE, assuming Theta0 = 0.9, in the 2x2x4 design.

The figures follow the number of individuals required in EMA's study (figure 15) that is considerably larger than the two other models. However, the probability of wrongly concluding about bioequivalence is within the acceptable error of 5%, allowed for the alpha error (type I).

In the FDA protocol (figure 16) - and the proposed model (figure 17), the similarity between the two is quite marked. For the FDA protocol, there is an acceptable alpha error up to CVintra values close to 11, and the intermedium error range allowed in this study - between 5% and 10% - ends in values close to 12. For larger CVintra values, the probability of concluding BE increases significantly.

In the "New model" (figure 17), the CVintra values range is between 10 and 16, corresponding to an error between 5% and 10%, and for higher values, it is wrongly concluded as bioequivalent.

Design 2x2x4, Theta0=0,85

Figure 18, 19 and 20 show the results for formulations that are assumed to differ by 15%, under the 2x2x4 design.

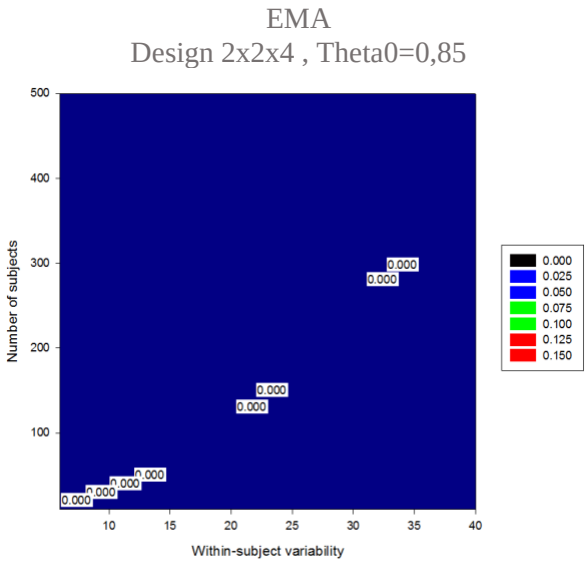

Figure 18 - EMA. Evaluation of the probability of wrongly concluding by BE, assuming Theta0=0.85, in the 2x2x4 design.

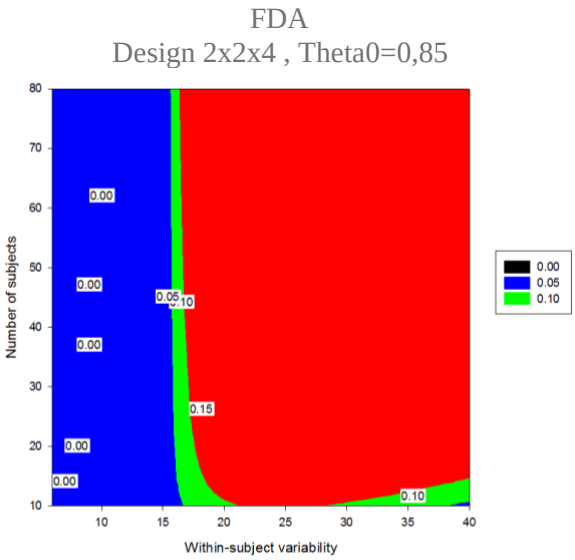

Figure 19 - FDA. Evaluation of the probability of wrongly concluding by BE, assuming Theta0=0.85, in the 2x2x4 design.

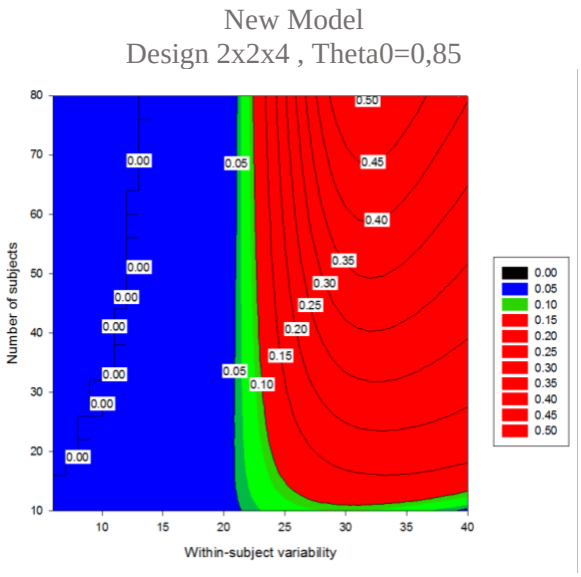

Figure 20 - New Model. Evaluation of the probability of wrongly concluding by BE, assuming Theta0 = 0.85, in the 2x2x4 design.

Following the conclusions on the previously accepted 10% difference, the current EMA protocol presents a very low probability (much less than 0.05) of wrongly conclude for bioequivalence in this scenario (figure 18).

However, the range of values marked as the acceptable error are on the 16% - 21% CV intra range, for the FDA model (figure 19) showing that above these values, there is a significant

probability of concluding Bioequivalence under this scenario. For the newly proposed criteria, the acceptable error boundary is between 21% - 40% (figure 20), although the number of volunteers in variable studies are over this range of values. This shows that, when compared to the FDA approach, a more conservative result is obtained.

Design 2x2x4, Theta0=0,80

In this phase of the study, figures 21, 22 and 23 will be observed, representing differences between formulations of 20%, assuming the 2x2x4 design.

EMA

Design 2x2x4 , Theta0=0,8

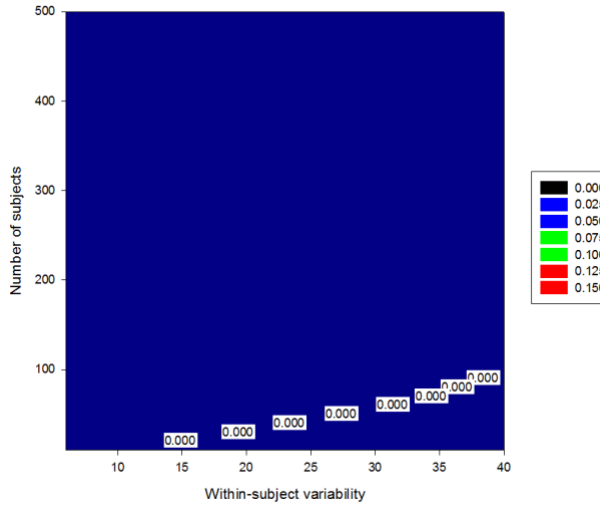

Figure 21 - EMA. Evaluation of the probability of wrongly concluding by BE, assuming Theta0=0.8, in the 2x2x4 design.

FDA

Design 2x2x4 , Theta0=0,8

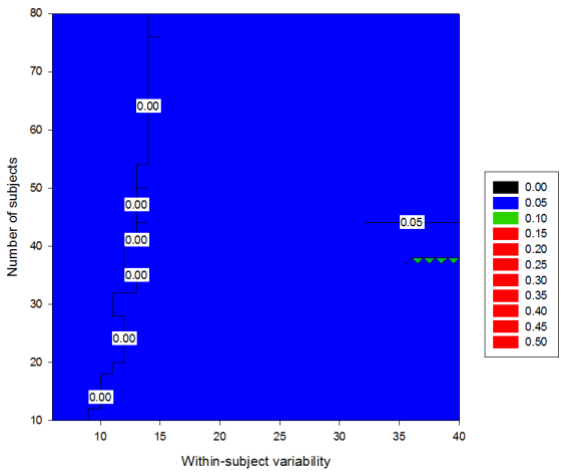

Figure 22 - FDA. Evaluation of the probability of wrongly concluding by BE, assuming Theta0=0.8, in the 2x2x4 design.

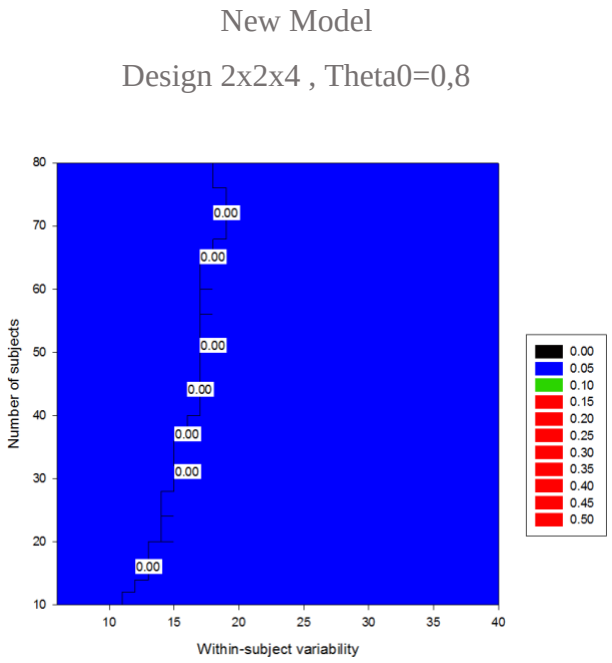

Figure 23 - New Model. Evaluation of the probability of wrongly concluding by BE, assuming Theta0=0.8, in the 2x2x4 design.

The analysis of the figures shown above allows us to conclude that when facing formulations whose difference can go up to 20%, the probability of concluding bioequivalence, for all models under study, is less than the acceptable alpha error of 5%.

Design 2x3x3, Theta0=0,90

The probability of wrongly concluding BE when it is assumed that the formulations are 10% different from each other, with a 2x3x3 study design, applying only to EMA's model and to the new model in study is shown below in figures 24 and 25.

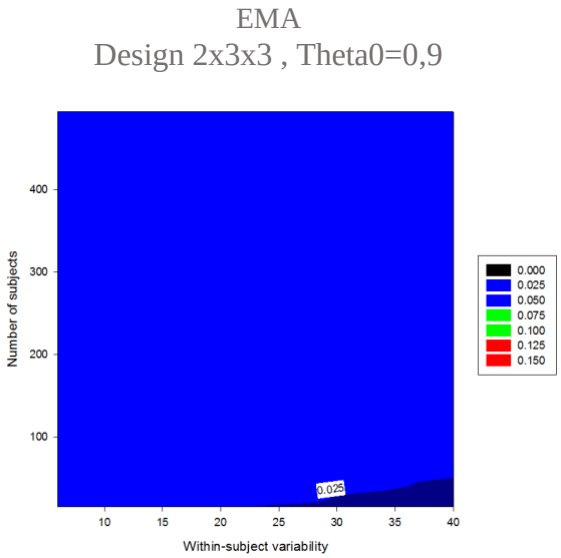

Figure 24 - EMA. Evaluation of the probability of wrongly concluding by BE, assuming Theta0=0.9, in the 2x3x3 design.

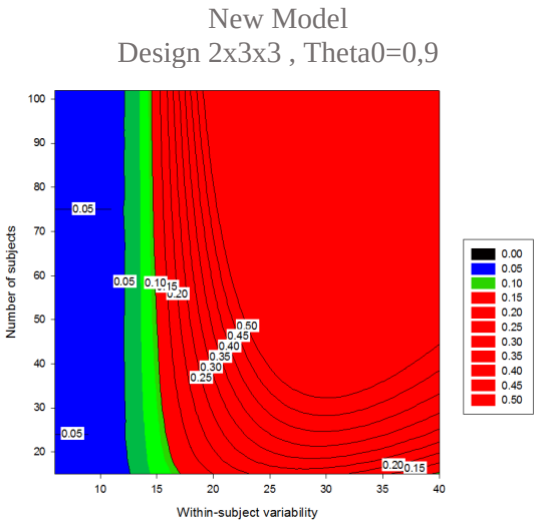

Figure 25 - New Model. Evaluation of the probability of wrongly concluding by BE, assuming Theta0=0.9, in the 2x3x3 design.

It is possible to observe that the EMA model (figure 24) rejects bioequivalence through an acceptable error of 5% as shown previously.

From the analysis of figure 25, referring to the new model, we can see an extension of the flexibility with which these NTDI drugs would be approved, with drug products with a 10% difference being possible to be considered bioequivalent for CVintra values above 13%.

Design 2x3x3, Theta0=0,85

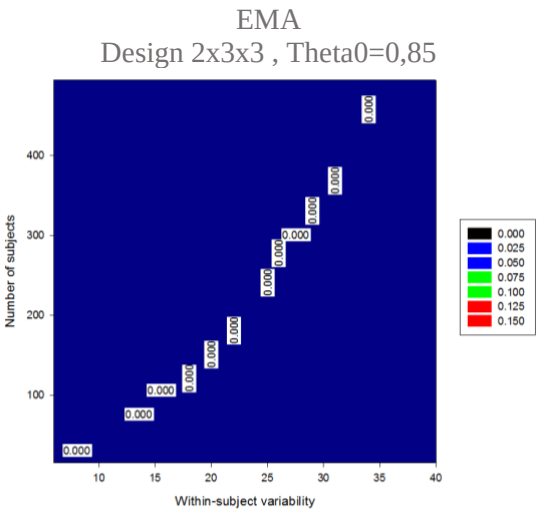

Figure 26 - EMA. Evaluation of the probability of wrongly concluding by BE, assuming Theta0=0.85, in the 2x3x3 design.

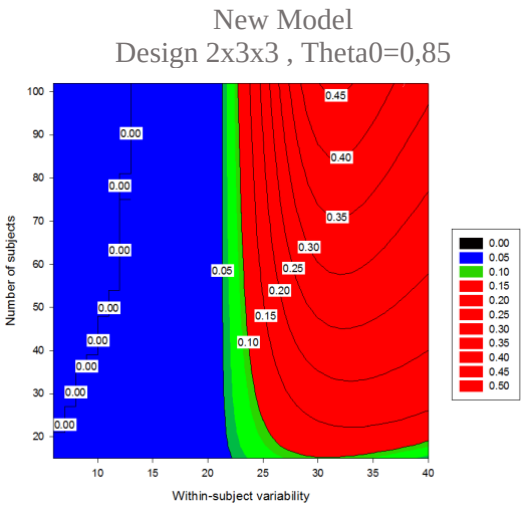

Figure 27 - New Model. Evaluation of the probability of wrongly concluding by BE, assuming Theta0=0.85, in the 2x3x3 design.

Through the analysis of figure 26 it is possible to understand that the outlook here is quite similar to that verified for formulations with differences of 10%. However, the probability of wrongly concluding by BE is, as expected, even lower.

Figure 27 shows again a different outlook, which is mostly independent on the number of volunteers used to demonstrate bioequivalence. In this case, it is shown that the chances to conclude bioequivalence between formulations differing more than 15% are only relevant for CVintra values above 22%.

Design 2x3x3, Theta0=0,80

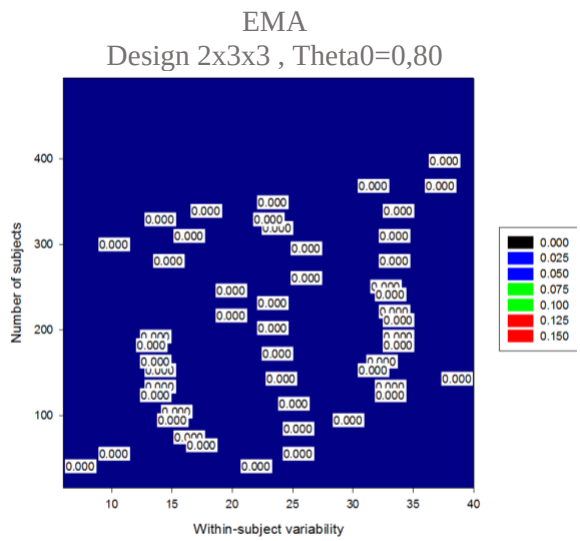

Figure 28 - EMA. Evaluation of the probability of wrongly concluding by BE, assuming Theta0=0.8, in the 2x3x3 design.

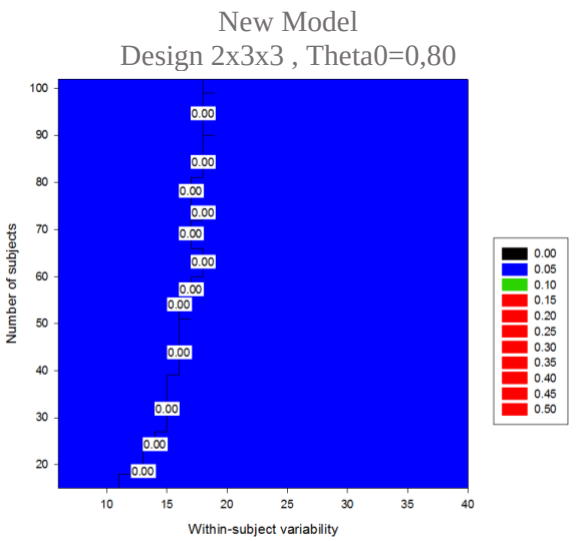

Figure 29 - New Model. Evaluation of the probability of wrongly concluding by BE, assuming Theta0=0.8, in the 2x3x3 design.

Figures 28 and 29 represent the probability of erroneously concluding for BE when formulations differ by 20% for the 2x3x3 design. These allow us to conclude that in both models, no matter the CVintra and number of subjects in the study, the conclusion regarding BE is always made with an acceptable error of not more than 5%, even though this error is lower for the EMA model.

Conclusion

In the first phase of the study, for the EMA model, we concluded that the number of individuals grows more than proportionally according to the variation coefficient when we want to prove bioequivalence for approval of NTDI drugs. These results allow us to conclude that the approval of NTDI drugs becomes problematic when it presents an intra-individual variation between 20-25% due to the excessive number of individuals that leads to high and unbearable costs and the greater possibility of abandonment during the study.

One way to deal with this problem is to use criteria with an interval extension based on CVintra, which is possible to observe in the FDA protocol and the proposed new model.

In the FDA model, that requires the use of the 2x2x4 design for quantifying both the test and reference CVintra, the opposite growth is observed when studying formulations equal to each other or with a difference of 5%.
For formulations that are the same, there is a growth in the number of individuals. However, these values appear to be less demanding when compared to the EMA protocol.
The results for formulations with a difference of 5% indicate an exponential decrease, less accentuated, which allows us to conclude that fewer individuals are needed to be in the study when we want to approve NTDI drugs.

When analyzing the new model under study, it shows to be apparently very similar to the FDA model but is more demanding in the number of individuals under investigation for lower coefficients of variation.

These results appear to be apparently positive for the FDA protocol and for the newly proposed model. However, we have to underline that the lower requirement in the number of volunteers needed in the study leads to the approval of formulations with progressively more significant differences, even if never exceeding 20%, a value considered as the "Golden-Standard" of BE. The same is not valid for EMA's model, which does not allow NTDI drugs' approval with differences more significant than 10% between them, assuming an alpha error (type I), which is 5%.

Bibliography

1. (CHMP). Guideline on the Investigation of Bioequivalence. EMA. 2010;Rev. 1:1–27.
2. Hu F, Ivanova A. Adaptative Designs. Encycl Biopharm Stat. 2004;1(1):1–6.
3. Al-Jazairi AS, Blhareth S, Eqtefan IS, Al-Suwayeh SA. Brand and generic medications: Are they interchangeable? Ann Saudi Med. 2008;28(1):33–41.
4. Yu LX, Bing VL. FDA Bioequivalence Standards [Internet]. 2014. 465 p. Available from: https://books.google.com/books?id=qjBwBAAAQBAJ&pgis=1
5. Minas R. Regulamentação Farmacêutica na garantia da qualidade , eficácia e segurança dos medicamentos genéricos. Ordem dos Farm [Internet]. 2014;52. Available from: http://www.ordemfarmaceuticos.pt/xFiles/scContentDeployer%7B_%7Dpt/docs/Docs335.pdf
6. Patterson SD, Jones B. Viewpoint: Observations on scaled average bioequivalence. Pharm Stat - Wiley Online Libr. 2011;(1):1–7.
7. Tamargo J, Le Heuzey JY, Mabo P. Narrow therapeutic index drugs: A clinical pharmacological consideration to flecainide. Eur J Clin Pharmacol. 2015;71(5):549–67.
8. INFARMED. DELIBERAÇÃO N.º 70/CD/2012. 2012;66:37–9.
9. Yu LX. Quality and Bioequivalence Standards for Narrow Therapeutic Index Drugs [Internet]. U.S. Food and Drug Administration. 2011. p. 37. Available from: https://www.fda.gov/files/drugs/published/Quality-and-Bioequivalence-Standards-for-Narrow-Therapeutic-Index-Drugs.pdf
10. Meredith P. Bioequivalence and Other Unresolved Issues in Generic Drug Substitution. Clinical Therapeutics. 2003.
11. Mei-Ling Chen, Vinod Shah, Rabindra Patnaik, Wallace Adams, Ajaz Hussain, Dale Conner, Mehul Mehta, Henry Malinowski, John Lazor, Shiew-Mei Huang, Don Hare, 1 Lawrence Lesko, Douglas Sporn and RW. Bioavailability and Bioequivalence: An FDA Regulatory Overview. Pharm Res. 18(3–4):1–7.
12. Detlew Labes [aut C, Helmut Schütz [aut], [aut] BL. Power and Sample Size for (Bio)Equivalence Studies [Internet]. 2019. Available from: https://cran.r-project.org/web/packages/PowerTOST/PowerTOST.pdf