Development of software for calculating the results of medicinal substances bioequivalence

N V Shalaginova¹, E A Mikhlyakova²,³, D N Gribkov⁴, A V Mikhailina⁵ and E V Lebedeva⁶

¹ Department of Applied Mathematics and Computer Science, Vyatka State University, Kirov, Russia
² Secondary school with in-depth study of individual subjects of Stulovo, Kirov, Russia
³ Graduate student of Vyatka State University, Kirov, Russia
⁴ Head of the Department of Informatics and Records Management, Orel State Institute of Culture, Orel, Russia
⁵ Head of Department of Economics and Management, Orel State Institute of Culture, Orel, Russia
⁶ Department of Algebra and Mathematical Methods in Economy of Orel State University named after I.S. Turgenev, Orel, Russia

E-mail: svkrotkrotova@yandex.ru

Abstract. International organizations have adopted and recognized drug standards, which should be guided by developers in the field of pharmacology. In practice, there is a problem: how to determine whether a reproduced drug sample really satisfies the world standard for its pharmacological action. It is required to implement an algorithm for checking and confirming the bioequivalence of drugs. The purpose of the study is to automate the verification of standards using an appropriate algorithm, i.e. to develop an application for calculating the results of medicinal substances bioequivalence. The methodology is based on the assumption that the identity in the sense of the created effect of the pharmacokinetic curves of the drug concentration in the blood versus time for the test drug and the standard means their therapeutic equivalence.

Research methods include Python programming, use of libraries for data visualization; Visual Studio is an application development environment. The desired algorithm and requirements for the developed application are formulated in the results of the study, based on the study of theoretical issues of determining the medicinal substances bioequivalence. Finally, conclusions about the efficiency of the software are made and options for its improvement are proposed.

1. Introduction

Nowadays, the development of medicine has achieved such success that the same disease can be cured in different ways, by taking different drugs with the same assigned effect. However, not all of these drugs are equally active. There are drug standards recognized by international organizations that should be guided by. The practical problem is to determine whether a reproduced drug sample really satisfies the world standard in its pharmacological action. In other words, an algorithm is needed to verify and confirm their bioequivalence [1]. A bioequivalence study is a type of clinical comprehensive study of a
medicinal product, which is produced to determine the equality of the action of the compared drugs according to the following parameters: the active substance absorption rate; the active substance excretion rate; the amount of the drug.

The key idea of this work is based on the assumption that the identity (or sufficient closeness) in the sense of the created pharmacological effect of the pharmacokinetic curves of the drug concentration in the blood versus time for the test drug and the standard means their therapeutic equivalence. The part of the administered dose of the explored drug that reaches the bloodstream is distributed in the body in accordance with the pharmacokinetics of the drug. A certain concentration profile of the drug is created in the blood and it can be directly measured. The drug also reaches the site of action, and the created concentration leads to a certain pharmacological effect in accordance with the pharmacodynamics of the drug. Equivalence is considered sequentially through all stages [2]. The main interest is the identity of the levels of the resulting effect. It is assumed that under certain conditions the "nonequivalence" of the compared drugs is reflected in the process of their absorption to a greater extent than in all other processes. However, the portion of the dose reaching the bloodstream cannot be directly measured, the blood concentration of the drug can be measured only. In addition, blood in this case is considered as a "gate" through which the drug must pass in order to reach the site of action.

The aim of this study is to develop software for calculating the results of medicinal substances bioequivalence.

To achieve this goal, the following tasks were set: to carry out a theoretical study of the problem of determining the medicinal substances bioequivalence; review the analogs among the software for calculating bioequivalence results; compile the required algorithm for calculating bioequivalence; formulate requirements for the developed application; implement an efficient application as required.

The object of the study is the medicinal substances bioequivalence.

2. Materials and methods

For a quantitative comparison of the absorption process of drugs, it is necessary to select indicators that characterize it in full. The main indicator for comparing the degree of absorption is the so-called bioavailability ratio - the ratio of the areas under the pharmacokinetic curves for the test drug and the standard [3].

The standard procedure is to measure concentration levels of the compared drugs in blood after receiving a single dose or after repeated dosing for each subject in the study. Periods of concentration measurement for different drugs of the same subject are separated by a sufficiently long time interval (5-6 drug half-lives) necessary to exclude drug interaction [4]. Measurements of concentrations after a single dose of the drug are made during at least three or four half-lives of the drug or stopped after the concentration drops below the registration limits [5].

Bioequivalence is verified using a non-compartmental approach [6]. The non-compartmental approach is based on calculating the statistical moments of the measured FK profile, which is regarded as an analogue of the statistical probability density. Bioequivalence values are assessed or calculated directly from the registered pharmacokinetic curves [7]. The rules regulate the choice of the main equivalence indicators, which are estimated directly from the recorded concentration-time curves.

Bioequivalence is checked according to the following algorithm:

Step 1. Based on the initial data of the measured concentrations of a substance in the blood of subjects, individual values of pharmacokinetic parameters are calculated after taking drugs T and R.

Step 2. The individual values of the relative bioavailability and the relative degree of absorption are calculated.

Step 3. The adequacy of the selection time of blood sampling from subjects is checked.

Step 4. For further statistical analysis of the parameters C_{\text{max}} and AUC, the logarithms of their values are taken.

Step 5. Multivariate analysis of variance is carried out. A statistical test ANOVA (Analyze of Variances) is used to check the bioequivalence of drugs.
The general procedure for analysis of variance: the set of observations is grouped by factor; the mean and variance are calculated for each group; the total variance is calculated for all groups of factors; the shares of the total variance due to intragroup and intergroup interactions are calculated; using a special criterion, it is determined how significant the differences between the observation groups are and whether the influence of the corresponding factors can be considered significant [6].

The analysis of the influence of factors on the resulting feature Y is carried out using the Fisher test, during which the value of the ratio of the corresponding variance of the factor to the variance of the error caused by random unregulated factors is analyzed. The calculated F value of each factor is compared with the table value for the significance level a (in this case, 0.05) and the number of degrees of freedom df. If the calculated value is greater than the table value, then the hypothesis is rejected. If the calculated value is less, the hypothesis is confirmed. If the null hypothesis about the influence of a particular factor is rejected, the statement about the presence of the main effect of the corresponding factor is accepted [8].

When the hypothesis of the interaction of factors is rejected, the statement is made that the influence of factor A manifests itself differently at different levels of factor B. In this case, the results of the analysis are invalidated, and the effect of factor A is checked separately at each level of factor B using one-way ANOVA.

At the end of the analysis of variance, the significance of the influence of factors on the studied value Y is checked. For this, the ratios of the variances of the factors to the sum of the variance of the factors interaction and the residual variance are calculated [9].

Step 6. Confidence intervals are calculated based on the results of analysis of variance.

Step 7. A conclusion is made about the medicinal substances bioequivalence and related indicators are calculated.

3. Results
At the beginning, the stage of the algorithm “Calculation of individual values of FK parameters” is performed. The assessment of the bioavailability of a medicinal substance is carried out by comparing the values of the pharmacokinetic parameters that are obtained during the analysis of the concentration-time curves for the investigated and compared drugs. The parameters C_{max} and T_{max} are estimated by out-of-model methods, that is, the highest values of C concentrations for each pharmacokinetic profile of the subjects and the corresponding values of the time of taking a sample T are taken.

C_{max,i} = \max(C_i), where Q is the i^{th} pharmacokinetic profile of the subject’s concentration.

T_{max,i} = T[\max(C_i)], where T[\max(C_i)] is the value of the time of taking a sample T at the maximum value of the C_{max} concentration. Next, the parameters AUC_{0-t} and AUC_{0-\infty} are calculated. To obtain the AUC_{0-t} parameter, the definite integral of the concentration-time curve of the subject must be calculated. The AUC_{0-\infty} parameter is calculated based on the elimination rate constant λ. The constant is found with linear regression using logarithmic transformations of the concentration value. From a certain point after the drug is administered to the subject, the concentration of the substance in the blood begins to decrease, as the drug is excreted from the body. This indicates the presence of decreasing linear regression. After calculating all individual pharmacokinetic parameters, the “Multivariate ANOVA” algorithm stage is carried out.

The null hypothesis H0 is put forward about the absence of differences between the compared drugs. It is assumed that the studied indicator (X_{ijk} – individual observation of the variable at j - subject after receiving the i - drug in the k sequence) can be presented as the sum of the total average (μ); parameter (σ_{ijk}) reflecting a reaction specific to a given j-subject; parameter τ_i, reflecting the effect of taking the i-drug; parameter π, characterizing the contribution associated with the l – measurement period; and ε_{ijk} - the random error of the model. It is assumed that the error has a normal distribution with zero mathematical expectation and variance DE. Also, a normal distribution of data, randomization of volunteers and the similarity of the variances of indicators for the compared drugs are required for the correct operation of the mathematical data model for the analysis of variance.
The hypothesis of the insufficient contribution of differences between drugs to the total variation of the data is tested using the F-test at a certain level of significance. However, this approach can be applied only on the assumption that there is no mutual influence of the drugs in the intervals of time of the corresponding measurements. If the time period between taking different drugs was not enough to completely eliminate any effect, this assumption is violated, and the results turn out to be incorrect. If this assumption is refused, the necessary mathematical modifications of the model are too complex. Another problem is the number of subjects. Even small differences can be statistically proven with a large sample size. With a small sample size, no significant differences can be found. To accommodate this, the "80/20" rule is introduced. This means that the study must be carried out so that the power of the statistical test for detecting 20% of statistically significant differences between drugs is not less than 80%. The calculated value of the F-test is compared with the tabular value for the two-sided test at the selected significance level. But this test only establishes the presence or absence of differences. Thus, regardless of the power of the test, it does not adequately verify the "approximate equivalence" hypothesis. Therefore, this approach is used only to test hypotheses about the statistical significance of the influence of various factors: differences between drugs (therapy), interindividuation differences (subjects), sequence of drug taking and study periods on the observed variation in data. And the residual variation obtained as a result of variance analysis is used when planning the required number of patients to be included in the study and calculating the corresponding confidence interval. For a common randomized crossover design, ANOVA includes the following factors: differences between drugs; interindividuation differences; the sequence of taking drugs; study periods.

Analyze of Variances and the criterion Student are parametric methods and assume a normal distribution of data. In case of violation of the distribution or other parameters, the logarithm of the indicators can be applied. The results after such a transformation can be interpreted as geometric means. In bioequivalence studies, the ratios of the means are often more interesting than the differences between them. Using a logarithmic transformation allows to go from the differences between the means of the converted data to the ratio of the corresponding means of the original data. The final stage of the statistical analysis of the compared drugs is the calculation and construction of confidence intervals based on the residual variations obtained in the course of multivariate analysis of variance. Also, at the end of the bioequivalence study, CVintra and the power of a statistical test to estimate the sample size of volunteers for the study can be calculated. The intra-individuation variability estimate (CVintra) is calculated based on the mean square of the error, which is calculated in analysis of variance. The power of a statistical test is the probability of rejecting a null hypothesis when it is not true. Moreover, the "powerful" statistical criterion is a criterion for which this probability is high. The power depends on the sample size; scatter; selection of a statistical test; differences between populations.

Let's review analogs for the developed software.

I version. "Bioequivalence 2.3". The software has the following features: analysis of samples from up to 70 volunteers; the number of blood samples from each volunteer at each stage - up to 29 inclusive; calculation of individual pharmacokinetic parameters based on concentrations and time of blood sampling; application of parametric analysis of variance ANOVA.

II version. "BIOEQV". The program was developed by I.B. Bondareva. It was used for the last 6 years for statistical processing of data from studies of bioequivalence of drugs. Features are full parametric and nonparametric statistics of indicators; information about individual and average differences; analysis of variations, construction of a confidence interval; construction of the distribution histograms of pairwise geometric means.

Thus, the following key features of the software for calculating bioequivalence results have been identified: calculation of basic pharmacokinetic parameters; calculation of the main indicators of mathematical statistics; multivariate analysis of variance; data visualization; calculation of confidence intervals [10].

The following stage of research is software development. Let’s describe the technical features and the justification of the choice of development methods. As a result of a theoretical study of bioequivalence and a review of analogs, the following requirements for the developed application were
formulated: the ability to obtain initial data from xls/xlsx files; calculation of all the necessary parameters of mathematical statistics based on the data obtained; carrying out multivariate analysis of variance, calculating confidence intervals; visualization of performed calculations in the form of tables and graphs; formation of a report on the results of bioequivalence. To implement the above described requirements, an analysis of the available development methods and tools was carried out.

The programming language Python was chosen to develop software as a windowed application. Python is a general-purpose programming language with many possibilities. There are many useful libraries to make development easier, such as NumPy for working with multidimensional arrays and matrices, PyQtGraph and MatPlotLib for data visualization. The PyQtGraph library was chosen to visualize the data in the form of graphs of different types. PyQt5, a set of Python libraries, was chosen to create the GUI for the application. PyQt is based on the Qt cross-platform library (used for GUI development in C++), extended for the Python language. The development environment was Visual Studio Code, a code editor that can be used not only on Windows and Mac OS X, but also on various Linux distributions [11]. At the same time, it is quite easy to learn, has a user-friendly interface and all the necessary functions for creating applications. If necessary, the problem can be resolved by installing additional extensions.

Application development. After formulation of requirements for the application and the choice of methods for its implementation, it was decided to develop software in the form of a windowed application with a user interface and visualization of calculations and data graphs. The main part of the application is the output of tables with calculations and the construction of graphs from the calculated data. The application has a sequential algorithm of actions: reading initial data from a file of xls /xlsx format; calculation of individual values of pharmacokinetic parameters; calculation of individual indicators of relative bioavailability; calculation of indicators of the adequacy of the time of blood sampling; logarithm of the main individual pharmacokinetic parameters; calculation of analysis of variance indicators; calculation of confidence intervals; report generation. The results and intermediate data of each step of the algorithm are displayed in a separate application window. All calculations are accompanied by visualization of intermediate data, which is also displayed in the application windows. Let’s consider in more detail each step of the general application algorithm.

Step I. Reading data. Data for analysis comes in Excel spreadsheet format. The application requires 3 xls/xlsx files; drug R concentration table; table of additional parameters for analysis of variance.

Additional parameters are the grouping of volunteers according to the following factors: drug; the sequence of taking the drug; period of taking the drug. Each factor has a value of 1 or 2, depending on the order of taking the drug T and R, respectively. Data reading is performed using the xldrd library. This library is designed to extract data from MS Excel spreadsheet files. This allows to receive data in a convenient format and work them. Using the obtained tabular values, the program builds graphs of the dependence of the each drug concentration in the blood from time for each subject.

Step II. Reading data. For the calculation of individual pharmacokinetic parameters, the IndividualFK class responds. The calcIndividFK() method accepts drug concentration data and mathematical statistics.

The MNK() method receives as input the concentration data of a particular subject for the entire time of the samples and calculates the linear regression indicators using the least squares method. The calculated indicators are displayed as a separate table for each drug T and R, respectively. The average pharmacokinetic profiles with outliers are plotted according to the indicators of mean and standard deviations.

Step III. Calculation of individual indicators of relative bioavailability. The BioAval class is responsible for the appropriate calculation and determination of the relative absorption rate. The bioAval() method receives arrays of drug T and R concentrations and individual FK parameters.

Step IV. Calculation of the adequacy of the sampling time indicators. The AdeqIndex class is responsible. The adeqIndex() method calculates the appropriate values. The graphs of standardized residuals in semilogarithmic coordinates are also built here. The linearReg() method calculates the residuals and returns an array of points used to plot the residuals relative to zero along the OX axis.
Step V. Logarithm of the main individual pharmacokinetic parameters. The LogFK class takes the logarithm of the FK parameters. The results are displayed in a table.

Step VI. Calculation of analysis of variance indicators. The ANOVA class gets logarithmic FK parameters and additional parameters from the file read earlier. The result is displayed in the form of a table. Two methods are used: for calculating confidence intervals and statistical indicators of test power, errors.

Step VII. Report generation. At the end, the program generates a report in Word format. It contains all the elements selected in the report settings. The DocumentCreate class is responsible for its formation.

4. Conclusion
At the testing stage, the results of the algorithm were evaluated. The software is a windowed application that displays all calculations for the user in the form of tables. At the start of the application work, the user selects data files and runs the algorithm. Each result according to the steps described above is displayed in a separate window.

So, at the initial stage of the study, the goal to develop software for calculating the results of bioequivalence of was set. To achieve the goal, a theoretical study on the bioequivalence of medicinal substances was carried out, an algorithm for calculating bioequivalence was proposed, requirements for the developed application were formulated, software for calculating the results of bioequivalence was developed and tested.

During the software development stage, a ready-to-use application was developed. This application can calculate the results of bioequivalence of medicinal substances. The program is a complete product and can be used to calculate the results of clinical bioequivalence studies.

The application can be improved in the following way: calculation additional indicators used in clinical bioequivalence studies; more detailed data visualization.

References
[1] Deng Y and Zhou X 2020 Methods to control the empirical type I error rate in average bioequivalence tests for highly variable drugs Statist. Meth. in Medical Research. 29(6) 1650-67
[2] Żebrowska M, Posch M and Magirr D 2015 Maximum type I error rate inflation from sample size reassessment when investigators are blind to treatment labels Statistics in Medicine 35
[3] Posch M, Klinglmueller F, König F and Miller F 2016 Estimation after blinded sample size reassessment Statist. Meth. in Medical Research 27
[4] Cicili B, Long T, Kim S and Cristofoletti R 2020 Assessing the impact of cystic fibrosis on the antipyretic response of ibuprofen in children: Physiologically-based modeling as a candle in the dark British J. of Clinical Pharm 86(11) 2247-55
[5] Nai C N, van Trung N, Thai T L, Hung T V and Tuan N D 2020 Simultaneous determination of amlodipine, losartan, and the losartan carboxylic acid active metabolite in human plasma by lc-ms/ms Intern. J. of Pharmac. Research. 12(4) 1981-90
[6] González-Rojano E, Marcotegui J, Laredo L, Gwaza L, Gordon J, Portolés A and García-Arieta A 2020 Chiral bioanalytical methods in bioequivalence studies of intravenous vs. oral formulations of ibuprofen Chirality 32(9) 1169-77
[7] Danielak D, Milanowski B, Wentowski K, Nogowska M, Kątny M, Rogowski P and Lulek J 2020 Physiologically based dissolution testing in a drug development Process – a case study of a successful application in a bioequivalence study of trazodone ER formulations under fed conditions AAPS PharmSciTech. 21(5)
[8] Reynolds K A, Pithadia D J, Lee E B, Liao W and Wu J J 2020 Safety and effectiveness of anti-tumor necrosis factor-alpha biosimilar agents in the treatment of psoriasis American J. of Clinical Dermatology 21(4) 483-91
[9] Mohammed A S, Woldekidan N A and Mohammed F A 2020 Knowledge, attitude, and practice
of pharmacy professionals on generic medicines in eastern ethiopia: A cross-sectional study

PLoS ONE 15(7)

[10] E V Soboleva and E V Kharunzheva 2019 J. Phys.: Conf. Ser. 1399 033094

[11] Sanders J and Kandrot E 2010 CUDA by Example: An Introduction to General-Purpose GPUProgramming (Addison-Wesley Professional) 312