

Использование языка программирования R для оценки биоэквивалентности лекарственных препаратов

Огурцов Андрей Борисович
ПАО НПЦ «Борщаговский ХФЗ»
Статистик

Аннотация

Статья содержит пример вычисления основных фармакокинетических параметров (AUC_{0-t} , C_{max}), построения соответствующих графиков и оценки биоэквивалентности двух препаратов при помощи языка R, являющегося примером свободного ПО (распространяется на условиях GNU General Public License). Сделаны выводы о целесообразности использования языка R в данной сфере в качестве альтернативы коммерческим программным продуктам для анализа данных.

Ключевые слова: R, ggplot2, shiny, биоэквивалентность, смешанная линейная модель.

Using R programming language to evaluate bioequivalence of drugs

Ogurtsov Andrey Borisovich
PJSC SIC «Borshchahivskiy CPP»
Statistician

Abstract

The article contains an example of calculation of the primary pharmacokinetic parameters (AUC_{0-t} , C_{max}), creating corresponding plots and evaluating bioequivalence of two drugs using R programming language (released under GNU General Public License). Conclusions about feasibility of using R language in this area as an alternative to commercial software products for data analysis were made.

Keywords: R, ggplot2, shiny, bioequivalence, mixed linear model.

Анализ полученных данных является важной частью любого клинического испытания, в том числе испытания с целью оценки биоэквивалентности. Список применяемого в данной области коммерческого программного обеспечения (ПО) достаточно обширен и включает как статистические пакеты общего назначения, такие как SAS, IBM SPSS Statistics, Statistica, Stata и SYSTAT, так и специализированные программы для анализа фармакокинетических данных. Примером специализированного ПО является WinNonlin – программа, разработанная компанией

Pharsight Corporation и позиционируемая разработчиками как «стандарт индустрии» [1].

Перечисленные программы достаточно разнообразны по своим возможностям и характеристикам: они могут иметь графический интерфейс или интерфейс командной строки (или одновременно оба); могут распространяться в виде единой программы или в виде отдельных компонентов, из которых собирается требуемая конфигурация; продаются в виде «вечной» лицензии на нужное количество рабочих мест или в виде годовой подписки. Объединяет их прежде всего высокая стоимость, что делает крайне актуальным поиск альтернативных средств в мире свободного ПО.

В данной статье описан опыт применения языка программирования R для решения типичных задач, возникающих при обработке фармакокинетических данных с целью оценки биоэквивалентности в исследованиях с перекрестным дизайном с двумя последовательностями приема препаратов и двумя периодами. Используемые подходы реализованы в виде веб-приложения, доступного по адресу <https://statist.shinyapps.io/bioeq/>. Исходный код также выложен в свободный доступ: <https://github.com/statist-bhfz/bioeq>.

Образец данных (без привязки к результатам исследований каких-либо конкретных препаратов) имеет вид, представленный в табл. 1. В таком же виде в файле с расширением .txt могут быть загружены и реальные данные для анализа (с сохранением тех же имен столбцов; значения в файле должны быть разделены символом табуляции, десятичный разделитель – запятая). Разумеется, из соображений конфиденциальности для этого следует использовать копию приложения, запущенную локально или на сервере предприятия.

Таблица 1. Формат исходных данных для анализа

subj	seq	prd	drug	time	conc
1	2	2	R	0	0
1	2	2	R	0,25	36,1
1	2	2	R	0,5	125

Переменные:

- subj – номер добровольца;
- seq – последовательность (1 – RT, 2 – TR);
- prd – период;
- drug – препарат (R – референтный, T – тестовый);
- time – время забора крови;
- conc – концентрация анализа.

Для корректного расчета AUC_{0-t} необходимо отметить имеющиеся нулевые концентрации для последней точки забора крови как отсутствующие значения (NA).

Все переменные, кроме концентрации анализа, должны быть закодированы как факторы для последующего анализа. Также потребуются создать новую переменную-идентификатор, включающую номер добровольца и принимаемый в данном периоде препарат:

```
data$seq<- factor(data$seq)
data$prd<- factor(data$prd)
data$drug<- factor(data$drug)
data$subj<- factor(data$subj)
data$subj2 <- paste(data$subj, data$drug, sep="")
data$subj2 <- factor(data$subj2,
levels=unique(data$subj2), ordered=TRUE)
```

Найти параметр C_{\max} для каждого добровольца и каждого препарата можно при помощи простого поиска наибольшего значения соответствующей переменной:

```
results<- data[data$time == 0, c(1:4)]
results$Cmax<- tapply(data$conc, data$subj2, FUN=max, na.rm=TRUE)
results$Cmax<- as.numeric(results$Cmax)
```

Для расчета AUC_{0-t} по методу трапеций применяем следующая функцию:

```
aucCalc<- function(conc, time) {
auc<- numeric(length(time)-1)
for (i in 2:length(time)) {
auc[i-1] <- 0.5*(conc[i]+conc[i-1])*
(time[i]-time[i-1])
}
return(sum(auc, na.rm=TRUE))
}
```

Использование этой пользовательской функции ничем не отличается от использования встроенных функций. Необходимо лишь предварительно извлечь вектор со значениями моментов времени забора крови из исходной таблицы с данными:

```
time<- unique(data$time)
results$auc<- tapply(data$conc, data$subj2, FUN=aucCalc, time=time)
results$auc<- as.numeric(results$auc)
```

Визуализация фармакокинетических кривых легко может быть выполнена при помощи графического пакета *ggplot2* [2] (рис. 1, 2, 3). Пример кода для создания рис. 1 представлен ниже.

```
ggplot(data, aes(x=time, y=conc, colour=subj2)) +
geom_line(size=0.6) + geom_point() +
guides(col=guide_legend(ncol=4)) +
theme(legend.title=element_blank()) +
labs(list(title="Фармакокинетические кривые для препаратов T и R",
x = "Время, ч", y = "Концентрация"))
```

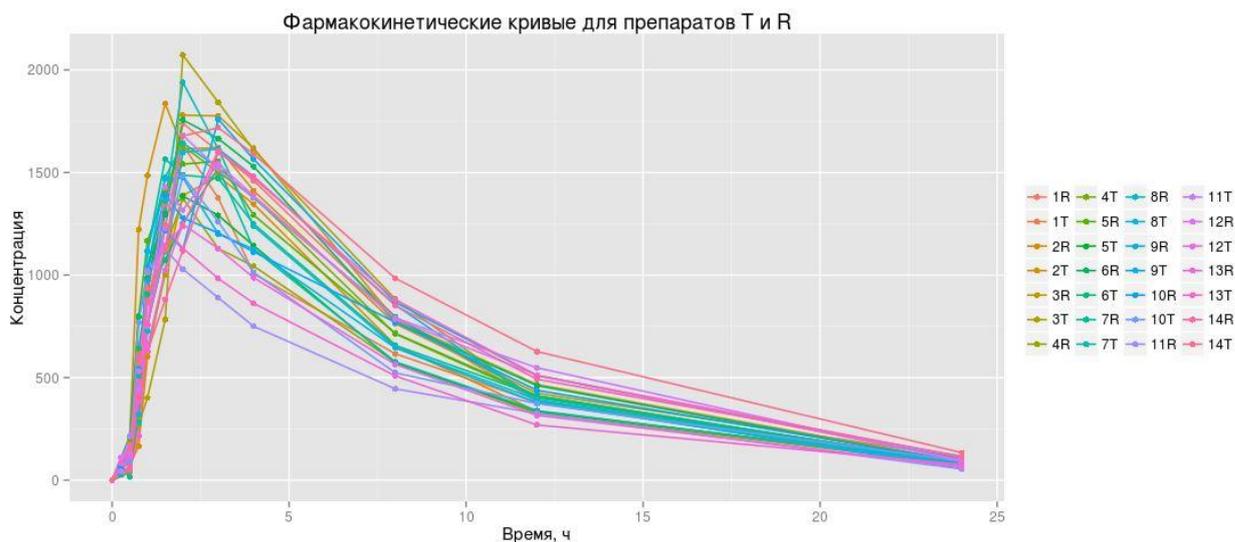


Рис. 1. Фармакокинетические кривые для препаратов Т и R

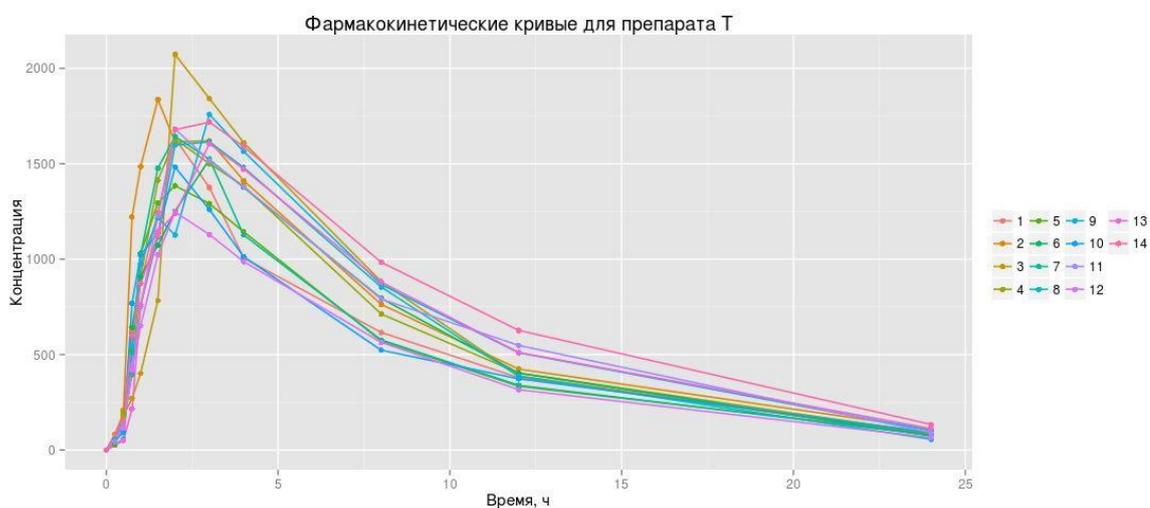


Рис. 2. Фармакокинетические кривые для препарата Т

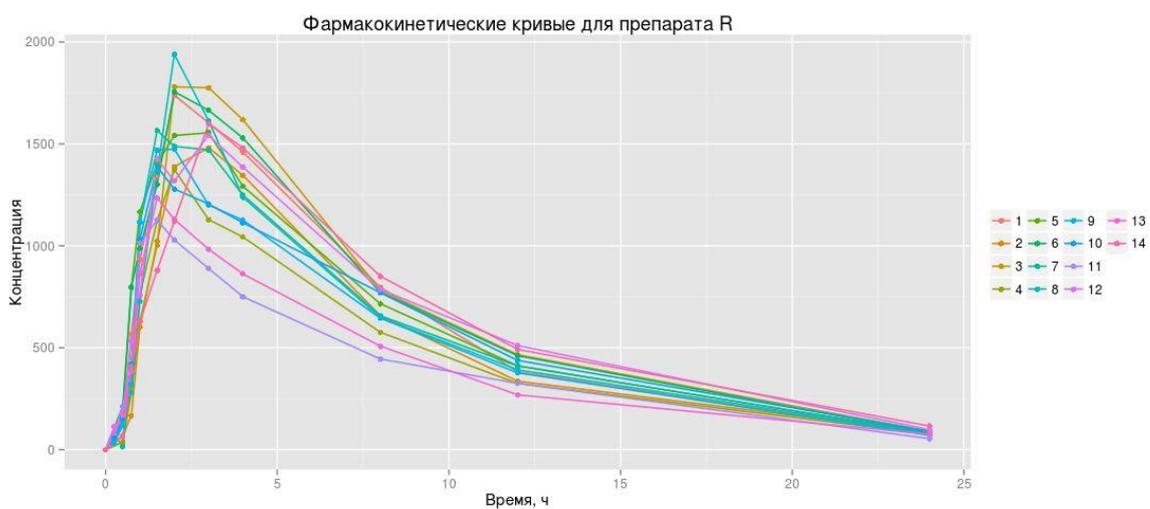


Рис. 3. Фармакокинетические кривые для препарата R

Статистический метод для оценки биоэквивалентности основывается на 90% доверительных интервалах для соотношения средних геометрических

(тестовый/референтный препарат) рассматриваемых параметров (AUC_{0-t} , C_{max}). Важно отметить, что в модели дисперсионного анализа, согласно действующим рекомендациям, предпочтительнее использовать фиксированные, а не случайные эффекты [3, 4], однако стандартным подходом для анализа повторных измерений является использование смешанных линейных моделей со случайными эффектами, заданными фактором «субъект». Этот вопрос обсуждается в [5], для дизайна с двумя периодами и двумя последовательностями оба подхода дают одинаковые результаты.

Для выполнения анализа с использованием смешанной линейной модели в R используем функцию *lme* из пакета *nlme* [6] (Табл. 2, 3):

```
model.Cmax<- lme(log(Cmax)~drug+prd+seq,
random=~1|subj, data=results)
model.auc<- lme(log(auc)~drug+prd+seq,
random=~1|subj, data=results)
```

Таблица 2. Результаты анализа логарифмически преобразованных данных C_{max}

	Value	Std.Error	DF	t-value	p-value
(Intercept)	7.36	0.05	12.00	142.87	0.00
drugT	0.06	0.05	12.00	1.27	0.23
prd2	-0.06	0.05	12.00	-1.31	0.21
seq2	0.00	0.06	12.00	0.03	0.98

Таблица 3. Результаты анализа логарифмически преобразованных данных AUC_{0-t}

	Value	Std.Error	DF	t-value	p-value
(Intercept)	9.18	0.12	12.00	74.61	0.00
drugT	-0.02	0.12	12.00	-0.15	0.88
prd2	0.04	0.12	12.00	0.30	0.77
seq2	0.13	0.12	12.00	1.07	0.31

Можно построить доверительные интервалы для каждого из коэффициентов моделей (нас будет интересовать коэффициент при *drugT*, поскольку именно он отражает различия между тестовым и референтным препаратом):

```
ci.Cmax<- intervals(model.Cmax, level=0.9, which="fixed")
ci.auc <- intervals(model.auc, level=0.9, which="fixed")
```

Для расчета соотношения T/R и получения 90% доверительных интервалов (ДИ) для исходных данных без логарифмирования служит следующий фрагмент кода (Табл. 4):

```

ci_Cmax<- exp(ci.Cmax$fixed[2, ])*100
ci_auc<- exp(ci.auc$fixed[2, ])*100

```

Таблица 4. Соотношение T/R и 90% доверительные интервалы

	Нижняя граница 90% ДИ, %	Соотношение T/R, %	Верхняя граница 90% ДИ, %
C _{max}	97.59	106.23	115.64
AUC _(0-t)	78.81	98.13	122.18

Представленные в статье результаты позволяют сделать выводы о целесообразности использования языка R и доступных для него дополнительных пакетов [2, 6, 7] в качестве инструмента для анализа данных в ходе обработки результатов клинических испытаний с целью оценки биоэквивалентности, а также для разработки прикладных программ, доступных конечному пользователю в удаленном режиме.

Библиографический список

1. Pharsight Products: WinNonlin // [Официальный сайт]. URL: http://pharsight.com/products/prod_winnonlin_trial.php (дата обращения: 13.03.2016).
2. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer, 2009.
3. Guideline on the Investigation of Bioequivalence. – CPMP/EWP/QWP/1401/98 Rev.1/Corr**. – European Medicines Agency, 2010.
4. Настанова 42-7.1:2014. Лікарські засоби. Дослідження біоеквівалентності. К.: Міністерство охорони здоров'я України, 2014.
5. Questions & Answers: Positions on specific questions addressed to the Pharmacokinetics Working Party (PKWP) – EMA/618604/2008 Rev.12. – European Medicines Agency, 2015.
6. Pinheiro J., Bates D., DebRoy S., Sarkar D. and R Core Team. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-125. – URL: <https://cran.r-project.org/web/packages/nlme/index.html> (дата обращения: 13.03.2016).
7. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. – URL: <http://www.r-project.org/> (дата обращения: 13.03.2016).