

© Криштафор А.А.

УДК 311.21; 214:57.03:51-76

**Криштафор А.А.**

## **СТАТИСТИЧЕСКИЙ АНАЛИЗ ДИНАМИЧЕСКИ ИЗМЕНЯЮЩИХСЯ БИОМЕДИЦИНСКИХ ПОКАЗАТЕЛЕЙ**

**ГУ «Днепропетровская медицинская академия МЗ Украины» (г. Днепропетровск)**

**chrishtaphor@gmail.com**

Данная работа является фрагментом НИР кафедры анестезиологии и интенсивной терапии ГУ «Днепропетровская медицинская академия МЗ Украины» «Разработка новых стратегий анестезиологического и периоперационного обеспечения в различных отраслях хирургии и вариантов интенсивной терапии критических и терминальных состояний с определением новых направлений заместительной и восстановительной терапии при использовании систем жизнеобеспечения», шифр ИН.Об.14.

**Вступление.** Организм человека представляет собой многокомпонентную систему, в которой каждый компонент и зависит от всех остальных, и сам оказывает влияние на все остальные компоненты. Вследствие большого количества элементов, от которых зависит та или иная функция организма, показатели, которые отражают состояние этой системы, являются интегральными. Это предопределяет сложность как их трактовки, так и их сравнение.

Живой организм характеризуется постоянным изменением функционального состояния его систем. Смыслом этих динамических изменений является адаптация организма к изменяющимся условиям внешней среды. Из этого следует, что показатели состояния любой системы организма являются актуальными только в момент их регистрации, т.к. отражают состояние системы именно в этот момент. Это необходимо учитывать при оценке состояния систем организма.

При проведении клинических исследований традиционное отношение к показателям, характеризующим функции систем организма, основывается на регистрации одномоментного значения. При этом регистрация большинства показателей проводится, как правило, не чаще одного раза в сутки, в то время, как ряд показателей системы гомеостаза за сутки претерпевает многократные колебания вокруг значения, которое, собственно, и отражает реальное состояние изучаемой системы. Поскольку частота таких колебаний превышает частоту измерений, в определенном смысле такие колебания могут быть названы высокочастотными.

С целью уменьшения влияния суточных колебаний активности систем организма, как правило, регистрация показателей проводится в одно и то же время суток. Однако, этот принцип не исключает влияния на уровень интегрального показателя множества других факторов, вызывающих высокочастотные колебания этого показателя.

Статистический анализ данных, имеющих высокую частоту колебаний между периодами регистрации,

может привести к неверным выводам. Выделяют два вида статистических ошибок: 1) заведомо ложные данные, результаты анализа которых также заведомо ложны; 2) ложные результаты, как следствие неправильного использования статистического инструментария [1, 4]. Поскольку значение, взятое в определенный момент времени, может отклоняться от средней величины показателя на значительную величину, то оно может считаться исходно ложным, а значит вызывать ошибку первого типа, которую невозможно идентифицировать, поскольку исходное значение истинно на момент регистрации.

Существует значительный ряд публикаций, описывающих правила проведения статистического анализа биомедицинских данных, наиболее частые ошибки и методы их устранения [3]. Так, одним из методов устранения ошибки, связанной с неоднородностью исходных данных является метод стандартизации [2, 5], однако он касается неоднородности состава групп, а не колебаний признака.

**Целью** данного исследования было изучение влияния высокочастотных колебаний признака на регистрируемое значение, разработка методики коррекции возможных искажений реальной величины показателя и обработки данных, имеющих высокий уровень искажения.

**Объект и методы исследования.** В исследовании принимали участие 10 больных после плановых хирургических вмешательств на органах брюшной полости. Исследование проводилось с использованием прикроватного монитора, обеспечивающего автоматическое измерение артериального давления неинвазивным способом через равные промежутки времени. Интервал между измерениями составлял 1 час. Для анализа выбраны значения систолического артериального давления, как одного из наиболее распространенных показателей, характеризующих состояние больного. Полученные значения обрабатывались с помощью параметрических методов статистики (ANOVA) с использованием табличного процессора программного комплекса LibreOffice.org, версия 4.3.

**Результаты исследований и их обсуждение.** Все полученные данные представлены в **таблице 1**. Анализ полученных значений проводился по следующей схеме: 1) сравнивались значения показателя, зарегистрированные в одно и то же время; 2) вычислялось значение средней величины, стандартной ошибки и ошибки среднего значения показателей, зарегистрированных в одно и то же время, у одного больного и у всей выборки больных.

Уровни систолического артериального давления у больных за весь период наблюдения (24 часа)

час	Номер больного									
	№1	№2	№3	№4	№5	№6	№7	№8	№9	№10
1	120	120	120	110	120	110	130	120	110	140
2	120	125	140	120	120	110	130	120	110	140
3	125	120	130	120	125	110	130	120	120	130
4	130	125	130	120	125	110	130	120	120	130
5	120	125	120	125	125	110	130	120	120	130
6	110	120	110	125	120	120	140	120	120	130
7	120	125	120	120	120	120	140	130	140	140
8	140	125	110	120	125	120	140	115	130	140
9	120	125	120	130	125	120	130	120	120	130
10	130	120	110	130	130	120	130	120	120	120
11	130	125	120	140	130	120	130	120	120	120
12	120	125	130	125	130	120	130	120	130	120
13	120	130	140	125	120	120	130	110	130	120
14	125	125	140	130	120	120	140	120	120	120
15	130	130	120	130	120	120	130	120	120	130
16	140	130	120	120	130	130	130	120	120	130
17	120	125	120	120	120	110	125	120	120	120
18	120	125	110	120	120	120	125	120	130	120
19	125	125	110	130	130	120	120	130	120	120
20	120	125	125	120	140	120	120	120	120	120
21	130	120	130	120	130	120	120	120	110	130
22	120	125	130	120	120	120	120	120	120	120
23	120	120	130	120	120	120	120	120	110	120
24	120	120	130	110	120	120	130	120	110	120

При проведении клинических исследований традиционно систолическое АД регистрируют в одно и то же время, например, в 8:00. Как следует из **табл. 1**, в это время показатель систолического АД у разных больных колебался от 110 мм рт.ст. до 140 мм рт.ст.. При этом, среднее значение показателя составило  $126,50 \pm 6,70$  ( $SD=10,81$ ) мм рт.ст. В то же время, например, у больного №1, у которого систолическое АД в 8:00 было максимальным среди зарегистрированных (140 мм рт.ст.), средний показатель за сутки был равен  $123,96 \pm 2,77$  ( $SD=6,91$ ) мм рт.ст., а у больного №3, у которого систолическое АД в то же время было минимальным среди зарегистрированных (110 мм рт.ст.), средний показатель за сутки был практически на том же уровне –  $123,54 \pm 3,84$  мм рт.ст. ( $SD=9,61$ ).

Но если бы исследователь измерил систолическое АД у тех же двоих больных спустя всего лишь час, он получил бы абсолютно одинаковые значения – 120 мм рт.ст.. При этом, изменения показателей у всех больных привели бы к тому, что среднее значение и ошибка среднего также существенно отличались бы –  $124,00 \pm 2,85$  мм рт.ст. ( $SD=4,59$ ).

В данном примере разница между полученными средними значениями с учетом доверительных интервалов незначительна, однако, если взять показатели, зарегистрированные в 16:00 и в 23:00, то получим  $127,00 \pm 4,18$  мм рт.ст. и  $120,00 \pm 2,92$  мм рт.ст. соответственно. При этом нижняя граница первого и верхняя граница второго доверительного интервала пересекаются всего лишь на 0,1 мм рт.ст., что при

анализе разных групп обычно трактуется как почти достоверная разница.

Уровень систолического артериального давления является результатом влияния различных факторов на деятельность сердечно-сосудистой системы, поэтому его колебания у каждого больного в каждый отдельный конкретный момент времени индивидуальны, а значит не могут быть спрогнозированы для группы больных.

Таким образом, можно сделать вывод, что быстроизменяющиеся показатели гомеостаза не должны учитываться по результатам единичных измерений. При малом объеме выборки колебания быстроизменяющегося показателя должны компенсироваться частотой его регистрации. При этом в качестве значения, соответствующего данному этапу исследования, очевидно, более корректно использовать среднее значение всех измерений, проведенных на данном этапе.

В связи с тем, что возможны значительные колебания исследуемого показателя даже в течение ограниченного времени его регистрации, при сравнении показателей разных групп или одного и того же показателя внутри группы, но на разных этапах, необходимо учитывать такой параметр как коэффициент изменчивости (вариации) показателя ( $C_v$ ) – отношение среднего квадратичного отклонения показателя к его среднему арифметическому значению, выраженное в процентах. Считается, что значение  $C_v$  ниже 5%

Таблица 2

Статистические показатели при коррекции исходных данных

Показатель	Исходная матрица данных	Матрица данных без предельных значений	Матрица данных с замещением предельных значений исходными средними
Среднее арифметическое	123,33	123,27	123,22
Стандартное отклонение	7,05	4,46	4,33
Доверительный интервал	2,82	1,78	1,73
Коэффициент изменчивости	5,72	3,62	3,51
Ошибка среднего арифметического	1,44	0,91	0,88

Таблица 3

Коэффициент изменчивости у больных без коррекции и с коррекцией исходных данных

Вид матрицы	Номер больного									
	№1	№2	№3	№4	№5	№6	№7	№8	№9	№10
Исходная	5,58	2,57	7,78	5,35	4,33	4,32	4,93	3,12	6,23	6,01
Без предельных значений	3,51	2,57	4,85	3,47	3,49	1,96	3,47	2,59	3,50	4,05
С замещением предельных значений средними	3,28	2,57	4,05	3,24	3,42	1,95	3,22	2,54	3,09	3,76

соответствует низкой изменчивости признака, от 5% до 10% – средней, а выше 10% – высокой.

Анализ таблицы исходных значений с помощью коэффициента изменчивости показателей показал следующее. Изменчивость признака у разных больных колебалась от 2,57% у больного №2 до 7,78% у больного №3. Коэффициент изменчивости в целом во всей группе составил 5,72, что может быть расценено как вариабельность средней выраженности.

Чем меньше изменчивость признака, тем выше качество статистического сравнения показателей разных групп или внутри группы на разных этапах. Для повышения качества анализа можно воспользоваться следующими приемами: 1) удалить из таблицы предельные значения; 2) заменить предельные значения, полученными с их участием, средними арифметическими величинами. Допустимость применения таких приемов исходит из того, что данные значения могли быть пропущенными, а значит к ним мог быть применен принцип замещения пропущенных значений. Результаты применения этих приемов отражены в **таблице 2**.

Анализ статистических показателей, полученных в результате коррекции исходной матрицы данных, позволяет сделать вывод, что даже простое удаление предельных значений в данном исследовании снижает степень изменчивости до низкой. Это касается не только показателей всей матрицы, но и данных по каждому больному в отдельности (**табл. 3**).

В случае, когда после замещения предельных значений коэффициент изменчивости признака всё ещё выше 5%, необходимо либо заместить следующие предельные значения, либо вообще исключить полученное у данного субъекта исследования среднее значение из группы, ввиду крайней изменчивости показателя, причины которой не известны или не могут быть устранены.

Таким образом, при статистической обработке данных исследований, включающих динамически изменяющиеся показатели, при учете таких показателей

на конкретном этапе необходимо в качестве регистрируемого параметра принимать среднюю арифметическую величину, полученную из серии замеров с коррекцией исходных значений в случае, если коэффициент изменчивости в этой серии превышает 5%. Применение такой методики позволит увеличить точность исследования за счет снижения влияния высокочастотных колебаний признака. Кроме того, анализ причин высокочастотных колебаний признака на отдельных этапах исследования может помочь в выявлении факторов, влияющих на течение заболевания и исход, которые ранее упускались из внимания.

**Выводы.**

Одномоментная регистрация показателей, частота изменения которых значительно превышает частоту регистрации, может быть источником ошибочных суждений о состоянии динамически изменяющейся системы.

Уменьшить ошибку измерения таких показателей можно за счет получения средней арифметической серии измерений на каждом этапе регистрации. При этом, чем больше измерений произведено, тем точнее средняя арифметическая отражает реальное состояние системы.

Параметры с высокой степенью изменчивости, определенной с помощью коэффициента изменчивости, должны исключаться из исследования ввиду того, что они вносят существенную ошибку в полученный результат.

Снизить степень изменчивости можно за счет исключения предельных значений или замещения их средней арифметической, полученной после исключения предельных значений.

**Перспективы дальнейших исследований.**

Дальнейшее исследование влияния высокочастотных флуктуаций признака на его регистрируемую на конкретном этапе величину может быть сосредоточено как на разработке методов выявления факторов, вызывающих эти флуктуации, так и на формулировке правил регистрации таких признаков.

### Литература

1. Леонов В.П. Ошибки статистического анализа биомедицинских данных / В.П. Леонов // Международный журнал медицинской практики. – 2007. – Вып. 2. – С.19-35.
2. Маймулов В.Г. Основы научно-литературной работы в медицине / В.Г. Маймулов, В.С. Лучкевич, А.П. Румянцев, В.В. Семенова. – СПб. : Изд-во “Специальная литература”, 1996. – изд. 2-е, испр. и доп. – 126 с.
3. Мараховский Ю.Х. Адекватность статистического анализа для получения научно обоснованных, достоверных результатов в исследованиях по клинической медицине / Ю.Х. Мараховский. – Код доступа [http://belmapo.by/downloads/gastroenterology/2009/stat/adekvatn\\_statich\\_analiza.pdf](http://belmapo.by/downloads/gastroenterology/2009/stat/adekvatn_statich_analiza.pdf).
4. Реброва О.Ю. Описание статистического анализа данных в оригинальных статьях. Типичные ошибки / О.Ю. Реброва // Медицинские технологии. Оценка и выбор. – 2011. – № 4. – Код доступа <http://www.hta-rus.ru/magazine-archive/102-pisanie-statisticheskogo-analiza-dannyih-v-originalnyih-statyah-ipichnyie-oshibki/>.
5. Glantz S.A. Biostatistics: how to detect, correct and prevent errors in the medical literature / S.A. Glantz // Circulation. – 1980. – Vol. 61. – P. 1-7.

УДК 311.21;:214:57.03:51-76

#### СТАТИСТИЧНИЙ АНАЛІЗ БІОМЕДИЧНИХ ПОКАЗНИКІВ, ЩО ДИНАМІЧНО ЗМІНЮЮТЬСЯ

Криштафор А.А.

**Резюме.** З метою визначення впливу добових коливань показника на достовірність його рівня, зареєстрованого в якості етапного показника клінічного дослідження, та розробки методів корекції, які дозволили б враховувати ці коливання, у 10 хворих протягом доби щогодинно реєстрували систолічний АТ. Виявилося, що у окремих хворих коливання АТ протягом доби були досить вираженими (коефіцієнт варіації становив до 8%), що може спотворювати результат при реєстрації в залежності від часу реєстрації. Запропонована методика корекції даних, що реєструються на етапі клінічного дослідження, шляхом багаторазової реєстрації показника з виключенням крайніх значень, або заміни їх на середні величини у випадку, значної варіабельності показника. Штучне зменшення варіабельності показника з реєстрацією в якості етапного показника середньої величини ознаки, дає можливість уникнути помилок в подальшому статистичному аналізі.

**Ключові слова:** біомедична статистика, клінічні дослідження, усунення помилок реєстрації показників.

УДК 311.21;:214:57.03:51-76

#### СТАТИСТИЧЕСКИЙ АНАЛИЗ ДИНАМИЧЕСКИ ИЗМЕНЯЮЩИХСЯ БИОМЕДИЧНИХ ПОКАЗАТЕЛЕЙ

Криштафор А.А.

**Резюме.** С целью определения влияния суточных колебаний показателя на достоверность его уровня, зарегистрированного в качестве этапного показателя клинического исследования, и разработки методов коррекции, которые позволили бы учитывать эти колебания, у 10 больных в течение суток ежедневно регистрировали систолическое АД. Оказалось, что у отдельных больных колебания АД в течение суток были достаточно выраженными (коэффициент вариации составил до 8%), что может исказить результат при регистрации в зависимости от времени регистрации. Предложена методика коррекции данных, регистрируемых на этапе клинического исследования, путем многократного измерения показателя с исключением крайних значений или заменой их на средние величины в случае, значительной вариабельности показателя. Искусственное уменьшение вариабельности показателя с регистрацией в качестве этапного показателя средней величины признака, дает возможность избежать ошибок в дальнейшем статистическом анализе.

**Ключевые слова:** биомедицинская статистика, клинические исследования, устранение ошибок регистрации показателей.

UDC 311.21;:214:57.03:51-76

#### STATISTICAL ANALYSIS OF DYNAMICALLY CHANGING BIOMEDICAL DATA

Krishtafor A.A.

**Abstract.** The human body is a multi-component system wherein each component depends on all the others, and itself has an effect on all other components. A living organism is characterized by constant changes in the functional status of its systems for a dynamic adaptation to changing environmental conditions. This implies that any indicators of its system are relevant only at the time of their registration, because reflects the system state at that moment. Usually in clinical studies indicators check no more than once a day, while a number of indicators of homeostasis per day undergoes repeated fluctuations around the value. Since the frequency of the oscillation exceeds the measurement frequency, in a certain sense, these fluctuations can be called high-frequency. The aim of this study was to investigate the effect of high-frequency oscillations of indicators for the registered value, development of methods of correction of possible distortions of the real value of the indicator and data with a high level of distortion.

**Material and methods.** The study involved 10 patients after elective surgery for abdominal organs. The study was conducted using a bedside monitor that provides automated non-invasive blood pressure measurement method at regular intervals. The values obtained were processed using parametric statistical methods (ANOVA) in a spreadsheet software package LibreOffice.org, ver. 4.3.

**Results and discussion.** In clinical studies traditionally systolic blood pressure were recorded at the same time, such as 8:00. In our study at this time systolic blood pressure in different patients ranged from 110 mmHg up to 140

mmHg. Thus, the average value of the index was  $126,50 \pm 6,70$  (SD = 10,81) mmHg. At the same time, for example, the patient No.1 had the systolic blood pressure maximum at 8:00 among registered (140 mmHg), the average per day =  $123,96 \pm 2,77$  (SD = 6,91) mmHg and the patient No.3 had systolic blood pressure at the same time minimal among registered (110 mmHg), the average per day was almost the same level –  $123,54 \pm 3,84$  mmHg (SD = 9,61). But if researchers would measure systolic blood pressure in the same patients after only an hour, he was received exactly the same values in both – 120 mmHg. Thus, changes indices of all patients would lead to the fact that the mean and error of mean values would also significantly different to –  $124,00 \pm 2,85$  mmHg (SD = 4,59). In this example, the difference between the mean values obtained with the confidence intervals is negligible, however, if we take the figures registered at 16:00, and at 23:00, we get  $127,00 \pm 4,18$  mm Hg and  $120,00 \pm 2,92$  mm Hg respectively. The lower limit of the first and the second upper limit of the confidence interval crossed by only 0,1 mm Hg and can be treated as significant difference. Thus, we can conclude that the rapidly changing parameters of homeostasis must not be taken into registered data as the results of single measurements and should be compensated frequency of its registration. And as the value corresponding to this phase of the study is more correct to use the average of all measurements taken at this stage.

Due to the fact that there are significant fluctuations of the rate even for a limited time its registration, if one compares the different groups, or the same rate within the group, but at various stages, we must consider this option as the coefficient of variation (Cv). The analysis of the reference values by means of coefficient of variability of indicators showed that variability of the sign at different patients fluctuated from 2,57% at the patient No.2 to 7,78% at the patient No.3. The variability coefficient in general in all group was 5,72 that can be regarded as variability of middle expressiveness. The variability of the sign is less, the quality of statistical comparison of indicators of different groups or in group at different stages is higher. For improvement of quality of the analysis it is possible to use the following receptions: 1) to remove extreme values from the table; 2) to replace extreme values, received with their participation, arithmetic averages. The admissibility of application of such receptions recognizes that these values could be passed, so the principle of replacement of the passed values could be applied to them.

The analysis of the statistics received as result of correction of the initial data, allows to conclusion, what even simple removal extreme values in this research reduces a degree of variability up to low. In case of when after replacement of extreme values the factor of variability of the sign still above 5%, is necessary to replace following extreme values or to exclude received at the given subject of research average value from group, in view of extreme variability of a parameter which reasons are not known or cannot be eliminated.

Application of such technique will allow to increase research accuracy due to decrease influence of high-frequency fluctuations of the indicator. Besides of that, the analysis of the reasons of high-frequency fluctuations of the indicator at separate investigation phases can help with identification of the factors influencing the course of a disease and an outcome which were missed from attention earlier.

*Conclusions.* One-stage registration of indicators, change which frequency considerably exceeds registration frequency, can be a source of wrong judgments about a condition of dynamically changing system. It is possible to reduce the error of measurement of such indicators due to receiving an arithmetic average of measurements series at each stage of registration. Thus, the more measurements it is performed, the arithmetic average is more exact reflects a real condition of system. Parameters with high degree of the variability defined by variability coefficient must be excluded from research in view of the fact that they bring an essential mistake in the received result. It is possible to lower variability degree at the expense of an exception of extreme values or replacement their by arithmetic average, received after exception of extreme values.

**Keywords:** biomedical statistics, clinical trials, elimination of registration error data.

*Рецензент – д.мед.н. Шкурупій Д.А.*

*Стаття надійшла 03.06.2015 р.*