

СОВРЕМЕННЫЕ ТЕНДЕНЦИИ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ДАННЫХ И ПРЕДСТАВЛЕНИЯ РЕЗУЛЬТАТОВ В ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЯХ

КУТИХИН А.Г.¹, ЮЖАЛИН А.Е.²

¹Федеральное государственное бюджетное научное учреждение «Научно-исследовательский институт комплексных проблем сердечно-сосудистых заболеваний», г. Кемерово, Россия

²Оксфордский институт радиационной онкологии, Оксфордский университет, г. Оксфорд, Соединенное Королевство Великобритании и Северной Ирландии

(по личному опыту исследовательской практики, а также материалам статьи Weissgerber TL, Milic NM, Winham SJ, Garovic VD. Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. PLoS Biol. 2015;13(4):e1002128. doi: 10.1371/journal.pbio.1002128)

LECTURE

HOW TO ANALYZE AND PRESENT EXPERIMENTAL DATA IN KEEPING WITH THE MODERN PARADIGM

ANTON G. KUTIKHIN¹, ARSENIY E. YUZHALIN²

¹Research Institute for Complex Issues of Cardiovascular Diseases (6, Sosnovy Boulevard, Kemerovo, 650002), Russian Federation

²Department of Oncology, Cancer Research UK and Medical Research Council Oxford Institute for Radiation Oncology, University of Oxford (Old Road Campus Research Building, Roosevelt Drive, Oxford OX3 7DQ), United Kingdom

(по личному опыту исследовательской практики, а также материалам статьи Weissgerber TL, Milic NM, Winham SJ, Garovic VD. Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. PLoS Biol. 2015;13(4):e1002128. doi: 10.1371/journal.pbio.1002128)

Резюме

В статье излагаются основные положения текущих трендов относительно статистической обработки данных и представления результатов в экспериментальных исследованиях: приоритет методов непараметрической статистики над параметрическими критериями, представление результатов в виде диаграмм рассеяния, а не в ви-

де столбиковых или линейных диаграмм, отказ от таблиц в пользу графиков, анализ результатов исследования в программе GraphPad Prism и принцип унификации рисунков внутри одной панели.

Ключевые слова: экспериментальное исследование, статистический анализ, представление результатов, непараметрические критерии, диаграмма рассеяния.

Abstract

Here we discuss the recent trends on statistical analysis and presentation of the experimental data. Among these are: 1) choice of non-

parametric instead of parametric tests for the statistical analysis; 2) presentation of the data as univariate scatterplots but not as bar or line graphs; 3) use of figures instead of tables; 4) use

◀ English

of GraphPad Prism for both statistical analysis and data presentation; 5) unification of the figures within one panel.

Keywords: experimental study, statistical analysis, data presentation, non-parametric tests, univariate scatterplot.

Одним из основных трендов биомедицинской науки в настоящее время является ужесточение требований к статистической обработке данных для максимально корректного представления результатов исследований. В то же время между клиническими и экспериментальными работами существуют значительные различия как в статистических методах, так и в способах представления результатов. Это обусловлено в первую очередь размерами выборок – клинические исследования включают от сотен до десятков тысяч пациентов, в то время как объекты исследования в экспериментальной практике исчисляются не более чем десятками.

Неизбежным следствием этого становится частая невозможность надежной оценки нормальности распределения в экспериментальных исследованиях, что требует обязательного использования исключительно непараметрических критериев для статистического анализа результатов. В биомедицинской непараметрической статистике классически используются критерий Краскела-Уоллиса (Kruskal-Wallis test), критерий Манна-Уитни (Mann-Whitney

test), критерий Фридмана (Friedman test), критерий Уилкоксона (Wilcoxon matched-pairs signed rank test) и критерий Данна (Dunn’s multiple comparisons test), каждый из которых применяется в зависимости от числа групп исследования, связанности выборок и необходимости внесения поправки на множественные сравнения (желающим детально ознакомиться с данным вопросом авторы рекомендуют переведенный на русский язык учебник Стентона Гланца «Медико-биологическая статистика» (Stanton A. Glantz. Primer of Biostatistics)).

В то же время относительно ограниченное количество объектов исследования дает возможность графически представить результаты измерений по каждому из них, что позволяет максимально полно отразить полученные экспериментальные данные. Стоит отметить, что практически во всех случаях, где это технически возможно, результаты экспериментов следует представлять не в таблицах, а в графиках, поскольку они значительно более наглядно и доступно передают информацию читателям журнала (**таблица 1, рисунок 1**).

Таблица 1. Пример представления результатов экспериментальных данных в виде таблицы

Концентрация криптонных молекул в сыворотке крови, нг/мл			
Меркуриане	Марсиане	Юпитериане	Нептуняне
219	227	314	588
245	234	435	954
257	366	476	876
259	488	392	824
277	564	246	788
323	601	288	901
345	876	197	534
349	545	165	478
367	935	400	700
370	967	371	766
378	767	354	634
411	645	176	601
443	618	289	556
503	534	299	487
544	520	438	965
612	456	165	980
654	765	414	545
765	442	307	453
Медиана и межквартильный размах			
368,5 (272,5 – 513,3)	554,5 [°] (452,5 – 765,5)	310,5 (233,8 – 403,5)	667,0 ^{* **} (542,3 – 882,3)

*p = 0,0004 при сравнении с меркурианами

**p = 0,0001 при сравнении с юпитерианами

°p = 0,0006 при сравнении с юпитерианами

Serum concentration of the crypton molecules, ng/mL			
Mercurians	Martians	Jupiterians	Neptunians
219	227	314	588
245	234	435	954
257	366	476	876
259	488	392	824
277	564	246	788
323	601	288	901
345	876	197	534
349	545	165	478
367	935	400	700
370	967	371	766
378	767	354	634
411	645	176	601
443	618	289	556
503	534	299	487
544	520	438	965
612	456	165	980
654	765	414	545
765	442	307	453
Median with interquartile range			
368.5 (272.5 – 513.3)	554.5 [°] (452.5 – 765.5)	310.5 (233.8 – 403.5)	667.0 ^{*,**} (542.3 – 882.3)

Таблица 1. Пример представления результатов экспериментальных данных в виде таблицы

Table 1. Experimental data presented as a table

*p = 0.0004 compared to mercurians

**p = 0.0001 compared to jupiterians

°p = 0.0006 compared to jupiterians

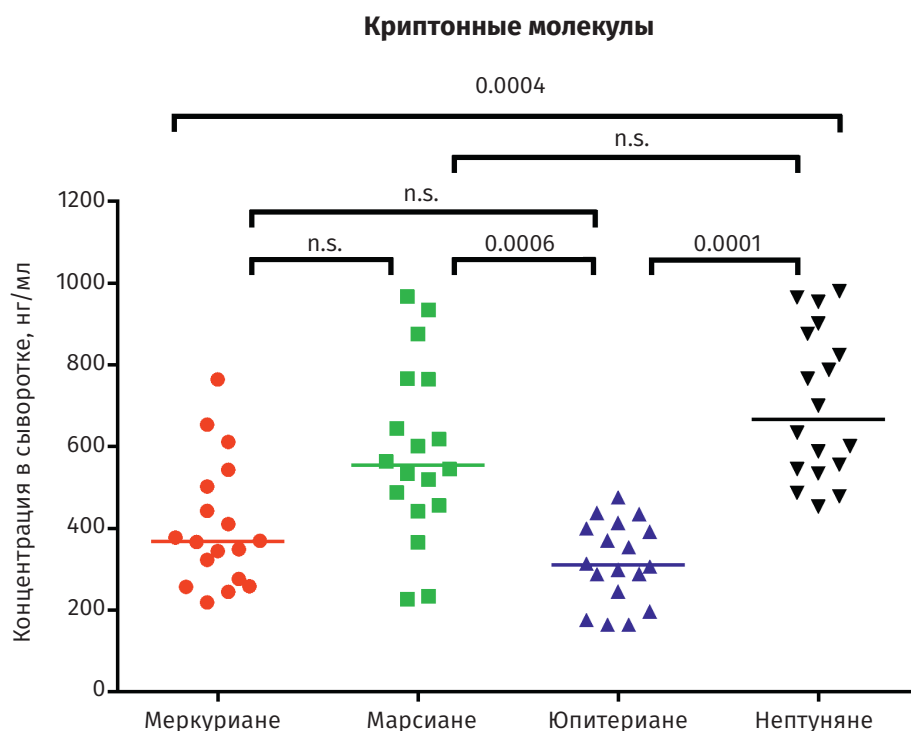


Рисунок 1. Пример представления экспериментальных данных в виде графика

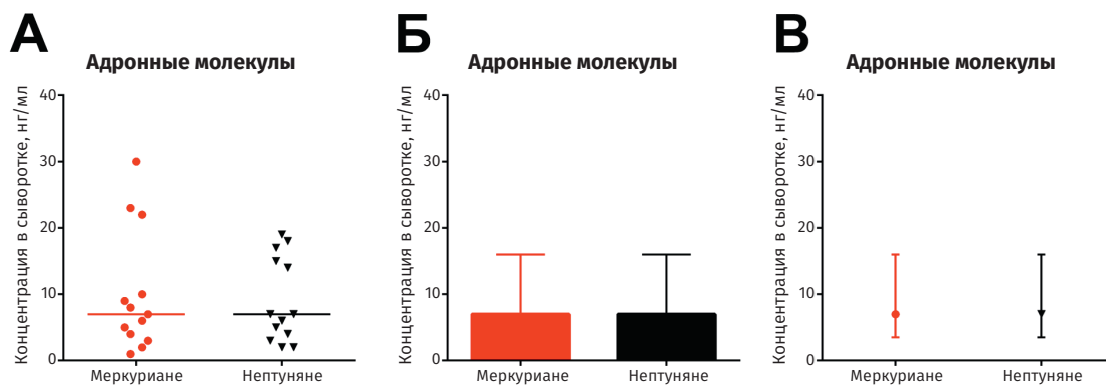
Figure 1. Experimental data presented as a figure

n.s. – нет статистически значимой разницы (от not significant). Значения вероятности отвергнуть верную нулевую гипотезу р представлены на графике численно.

n.s. is for not significant, p values are presented in a numerical manner.

Рисунок 2. Сравнение представления результатов по одному и тому же массиву данных в виде диаграммы рассеяния, столбиковой и линейной диаграмм. Как столбиковая, так и линейная диаграммы не позволяют читателям оценить распределение значений в целом и выпадающие значения в частности.

Figure 2. Univariate scatterplot compared to bar and line graphs. Either bar or line graphs do not allow the proper evaluation of the data distribution and outliers.



В отличие от таблиц, графики позволяют достаточно быстро и точно сравнить массивы данных между собой, в том числе учитывая статистическую значимость различий. Как правило, таблицы используют лишь для перечисления крупных массивов наименований и подсчетов (к примеру, результатов масс-спектрометрии или транскриптомного анализа) с последующим вынесением наиболее важных из них на график.

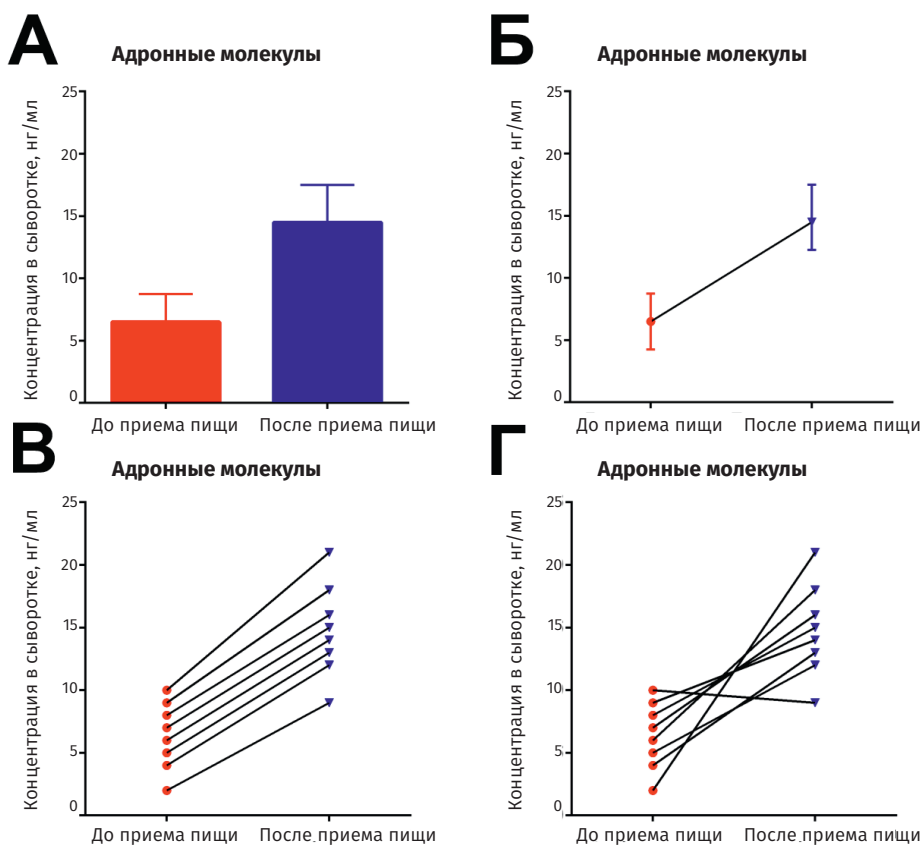
Чрезвычайно важным моментом в построении графиков по результатам экспериментов является вынесение результатов по каждому из объектов исследования на график, что позволяет сделать лишь диаграмма рассеяния (точеч-

ная диаграмма, scatterplot). В отличие от столбиковых и линейных диаграмм (bar graph и line graph), диаграммы рассеяния позволяют читателям в полной мере оценить характеристики выборки, в особенности значения, лежащие намного выше или ниже основного массива данных (выпадающие значения, outliers). Пример того, как столбиковая или линейная диаграмма может ввести читателей в заблуждение касательно природы выборки, представлен на **рисунке 2**.

Кроме того, диаграммы рассеяния идеально подходят для представления результатов повторных измерений, чего нельзя сказать о столбиковых или линейных диаграммах (**рисунк 3**).

Рисунок 2. Сравнение представления результатов по одному и тому же массиву данных в виде диаграммы рассеяния, столбиковой и линейной диаграмм. Как столбиковая, так и линейная диаграммы не позволяют читателям оценить распределение значений в целом и выпадающие значения в частности.

Figure 2. Univariate scatterplot compared to bar and line graphs. Either bar or line graphs do not allow the proper evaluation of the data distribution and outliers.



Таким образом, диаграммы рассеяния обеспечивают наиболее полное представление полученных результатов и позволяют их максимально объективную оценку в сравнении со столбиковыми и линейными диаграммами.

Следует также отметить, что для представления результатов экспериментальных, клинических и эпидемиологических исследований, как правило, применяются разные программы статистической обработки. Если последние обычно требуют специализированных программ (к примеру, Statistica или SPSS), то экспериментальные и клинические работы следует анализировать в программе GraphPad Prism, поскольку такой подход общепризнан в мировой научно-исследовательской практике. GraphPad Prism предоставляет как полный спектр статистических методов для обработки данных, так и все необходимые возможности для их надлежащего графического представления. Вследствие того, что аналогичные программы уступают GraphPad Prism в одном или в обоих указанных компонентах, не будет преувеличением утверждение о том, что не менее 98% экспериментальных работ в мировой науке используют для представления результатов именно эту программу.

При представлении результатов чрезвычайно важно помнить, что все непосредственно сравниваемые друг с другом фотографии внутри одной панели должны быть полностью сопоставимы по масштабу и иметь одинаковые шкалы. Крайне желательно, чтобы рисунки внутри одной панели были выполнены с использованием одних и тех же принципов и обозначений, а также в одной и той же цветовой гамме.

Вкратце основные правила статистического анализа и представления результатов для экспериментаторов можно изложить в следующей логической последовательности:

1. Перед анализом результатов следует убедиться в том, что Ваше исследование – экспе-

риментальное, а не клиническое или эпидемиологическое.

2. Далее необходимо оценить объем своих выборок и принять решение о том, следует ли анализировать тип распределения. Поскольку консенсуса по поводу порогового значения объема выборки для решения этого вопроса не существует, авторы данной лекции рекомендуют руководствоваться следующим эмпирическим правилом: не оценивать тип распределения и всегда использовать непараметрические статистические критерии. В экспериментальных работах это никогда не считается однозначной ошибкой. Кроме того, если хотя бы для одной из сравниваемых выборок нормальное распределение не характерно, это также требует применения именно методов непараметрической статистики.

3. Следует помнить о том, что статистический анализ и представление результатов экспериментальных работ должно осуществляться с использованием программы GraphPad Prism.

4. Необходимо максимально отказаться от представления результатов в виде таблиц в пользу рисунков. В виде таблиц допустимо приводить лишь масштабные массивы данных, но наиболее важные моменты при их анализе также следует представлять в виде рисунков.

5. При конструировании панелей нужно убедиться, что все непосредственно сравниваемые друг с другом фотографии внутри одной панели имеют одинаковый масштаб с обозначением шкал, а рисунки внутри одной панели выполнены с использованием одних и тех же принципов и обозначений, в одной и той же цветовой гамме.

В завершение стоит отметить, что ни в коем случае не следует недооценивать важность как правильно выполненного статистического анализа, так и максимально полного и объективного представления собственных данных. Оба этих момента являются критичными при анализе любого исследования и оцениваются рецензентами в первую очередь.●

Сведения об авторах

Кутихин Антон Геннадьевич, кандидат медицинских наук, старший научный сотрудник лаборатории геномной медицины отдела экспериментальной и клинической кардиологии ФГБНУ «Научно-исследовательский институт комплексных проблем сердечно-сосудистых заболеваний», г. Кемерово, Россия.

Вклад: написание лекции.

Южалин Арсений Евгеньевич, аспирант Оксфордского института радиационной онкологии, Оксфорд, Соединенное Королевство Великобритании и Северной Ирландии.

Вклад: написание лекции.

Корреспонденцию адресовать:

Кутихин Антон Геннадьевич
650002, г. Кемерово, ул. Сосновый бульвар, 6
E-mail: antonkutikhin@gmail.com

Authors

Dr. Anton G. Kutikhin, MD, PhD, Laboratory for Genomic Medicine, Division of Experimental and Clinical Cardiology, Research Institute for Complex Issues of Cardiovascular Diseases, Kemerovo, Russian Federation
Contribution: wrote the lecture.

Mr. Arseniy E. Yuzhalin, MSc (Res), PhD Student, Department of Oncology, Cancer Research UK and Medical Research Council Oxford Institute for Radiation Oncology, University of Oxford, Oxford, United Kingdom
Contribution: wrote the lecture.

Acknowledgements: There was no funding for this article.

Corresponding author:

Prof. Valeriy I. Podoluzhnyi,
Voroshilova Street 22a, Kemerovo, 650056,
Russian Federation
E-mail: pvi2011@mail.ru

Статья поступила: 11.01.17г.

Принята в печать: 26. 01.17г.