

## Применение диаграмм Венна и теории вероятностей в клинических исследованиях

Н.М. Буланов<sup>1,✉</sup>, О.Б. Блюсс<sup>1,2</sup>, Д.Б. Мунблит<sup>1,3</sup>, Т.В. Назаренко<sup>4</sup>, Д.В. Бутнару<sup>1</sup>,  
М.Ю. Надинская<sup>1</sup>, А.А. Заикин<sup>1,4</sup>

<sup>1</sup> ФГАОУ ВО «Первый Московский государственный медицинский университет им. И.М. Сеченова»

Минздрава России (Сеченовский Университет)

ул. Трубецкая, д. 8, стр. 2, г. Москва, 119991, Россия

<sup>2</sup> Университет Хартфордшира

Колледж-лейн, Хатфилд, AL10 9AB, Великобритания

<sup>3</sup> Имперский колледж Лондона

Экзибишн-роуд, Южный Кенсингтон, Лондон, SW7 2BU, Великобритания

<sup>4</sup> Университетский колледж Лондона

Гоуэр-стрит, Лондон, WC1E 6BT, Великобритания

### Аннотация

В настоящее время статистика играет ключевую роль при проведении медицинских исследований, объектом ее изучения являются накопление и обобщение данных, а также интерпретация взаимосвязей между экспериментальными данными. Данная статья открывает вводный цикл, посвященный проблемам биомедицинской статистики, и знакомит читателей с диаграммами Венна, а также с базовыми концепциями теории вероятностей и теории множеств, понимание которых необходимо для освоения принципов описательной статистики и статистики вывода. В статье представлено применение диаграмм Венна в современных клинических исследованиях, а также определены пространство элементарных событий, базовые операции с множествами (объединение, пересечение) и их применение в контексте классической теории вероятностей. Все примеры проиллюстрированы диаграммами Венна.

**Ключевые слова:** статистика; теория множеств; вероятность; диаграмма Венна

**Рубрики MeSH:**

КЛИНИЧЕСКИЕ ИССЛЕДОВАНИЯ КАК ТЕМА

ТЕОРИЯ ВЕРОЯТНОСТИ

СТАТИСТИКА КАК ТЕМА

**Для цитирования:** Буланов Н.М., Блюсс О.Б., Мунблит Д.Б., Назаренко Т.В., Бутнару Д.В., Надинская М.Ю., Заикин А.А. Применение диаграмм Венна и теории вероятностей в клинических исследованиях. Сеченовский вестник. 2020; 11(4): 5–14. <https://doi.org/10.47093/2218-7332.2020.11.4.5-14>

**КОНТАКТНАЯ ИНФОРМАЦИЯ:**

**Буланов Николай Михайлович**, канд. мед. наук, ассистент кафедры внутренних, профессиональных болезней и ревматологии ФГАОУ ВО «Первый МГМУ им. И.М. Сеченова» Минздрава России (Сеченовский Университет)

**Адрес:** ул. Трубецкая, д. 8, стр. 2, г. Москва, 119991, Россия

**Тел.:** +7 (919) 100-22-79

**E-mail:** nmbulanov@gmail.com

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

**Финансирование.** Исследование не имело спонсорской поддержки (собственные ресурсы).

**Статья поступила в редакцию:** 01.10.2020

**Статья принята к печати:** 26.10.2020

**Дата публикации:** 28.12.2020

## Venn diagrams and probability in clinical research

Nikolay M. Bulanov<sup>1,✉</sup>, Oleg B. Blyuss<sup>1,2</sup>, Daniil B. Munblit<sup>1,3</sup>, Tatiana V. Nazarenko<sup>4</sup>,  
Denis V. Butnaru<sup>1</sup>, Maria Yu. Nadinskaia<sup>1</sup>, Alexey A. Zaikin<sup>1,4</sup>

<sup>1</sup> *Sechenov First Moscow State Medical University (Sechenov University)  
8/2, Trubetskaya str., Moscow, 119991, Russia*

<sup>2</sup> *University of Hertfordshire  
College Lane, Hatfield, AL10 9AB, United Kingdom*

<sup>3</sup> *Imperial College London  
Exhibition Rd, South Kensington, London, SW7 2BU, United Kingdom*

<sup>4</sup> *University College London  
Gower Street, London, WC1E 6BT, United Kingdom*

### Abstract

Statistics is the science dealing with the collecting, summarizing and interpreting of associations in research data, and has a leading role in medical research. This article is an introductory publication in a series devoted to biomedical statistics. The aim of this article is to acquaint the readers with the basic concepts of Venn diagrams, probability and set theory, which are required to further understand descriptive and inferential statistics. First, we discuss the applications of Venn diagrams in current clinical research. Then we discuss the definitions of sample space, events, basic set operations (union and intersection) and their implementation in the classical approach to probability theory. All examples are introduced with Venn diagrams to illustrate the cases.

**Keywords:** statistics; set theory; probability; Venn diagram

#### MeSH terms:

CLINICAL STUDIES AS TOPIC

PROBABILITY THEORY

STATISTICS AS TOPIC

**For citation:** Bulanov N.M., Blyuss O.B., Munblit D.B., Nazarenko T.V., Butnaru D.V., Nadinskaia M.Yu., Zaikin A.A. Venn diagrams and probability in clinical research. *Sechenov Medical Journal*. 2020; 11(4): 5–14. <https://doi.org/10.47093/2218-7332.2020.11.4.5-14>

#### CONTACT INFORMATION:

**Nikolay M. Bulanov**, Cand. of Sci. (Medicine), Assistant Professor, Department of Internal, Occupational Diseases and Rheumatology, Sechenov First Moscow State Medical University (Sechenov University)

**Address:** 8/2, Trubetskaya str., Moscow, 119991, Russia

**Tel.:** +7 (919) 100-22-79

**E-mail:** nmbulanov@gmail.com

**Conflict of interests.** The authors declare that there is no conflict of interests.

**Financial support.** The study was not sponsored (own resources).

**The article received:** 01.10.2020

**The article approved for publication:** 26.10.2020

**Date of publication:** 28.12.2020

Статистика — это наука, объектом изучения которой являются сбор, обобщение, представление и интерпретация данных для оценки силы ассоциаций и проверки гипотез [1]. В настоящее время статистические методы играют ключевую роль в медицинских и биологических исследованиях. Поэтому понимание и применение базовых подходов к анализу данных являются необходимыми навыками для всех врачей и исследователей. Однако большинство специалистов в области здравоохранения в России не имеют практически никакой математической подготовки и обладают крайне ограниченным пониманием статистики, что зачастую приводит к неверной и некритичной интерпретации публикуемых медицинских данных, а также неспособности организовать и провести собственные исследования и проанализировать их результаты должным образом. В связи с этим мы решили опубликовать цикл статей, составляющих введение в курс биомедицинской статистики.

Этот курс изначально был разработан для обучающихся по программам магистратуры и PhD Университетского колледжа Лондона, а позднее адаптирован для студентов и обучающихся по программам постдипломного образования Сеченовского Университета. Основная цель курса — практическое применение полученных знаний, поэтому основное внимание в статьях будет уделено наиболее важным математическим концепциям и статистическим методам, которые необходимы для критического понимания современных клинических исследований, а также подготовки публикаций, соответствующих требованиям высокорейтинговых медицинских журналов.

Мы настоятельно рекомендуем читателям научиться самостоятельно представлять полученные данные в адекватной форме, использовать базовые статистические тесты и интерпретировать полученные результаты без помощи профессионального статистика. Однако мы также включили в этот курс описание более сложных статистических методов, таких как алгоритмы машинного обучения, для того чтобы познакомить читателей с возможными областями их применения. Еще одной целью этого курса является обеспечение эффективного сотрудничества между врачами и статистиками для разработки рационального дизайна клинических исследований, формулирования реалистичных целей, выбора оптимальных подходов и адекватной интерпретации результатов исследования. Все методы, обсуждаемые в статьях, могут быть использованы в любых статистических программах, как коммерческих (SPSS, Stata, SAS, MatLab и др.), так и бесплатных (R).

Материал подготовлен с учетом отсутствия математической подготовки у большинства читателей. Однако в курсе будут в упрощенном виде представлены некоторые математические концепции и формулы,

поскольку, с точки зрения авторов, это наиболее подходящий способ представления идей, лежащих в основе методов статистического анализа данных. Курс представлен на русском и английском языках, чтобы познакомить читателей с наиболее часто используемыми в научных публикациях терминами.

Первая статья посвящена основам теории множеств и теории вероятностей и проиллюстрирована примерами с использованием диаграмм Венна.

## ДИАГРАММЫ ВЕННА

Диаграммы Венна, также известные как диаграммы множеств, часто используют для иллюстрации логических взаимосвязей между двумя и более наборами данных в публикациях, освещающих результаты клинических и эпидемиологических исследований. Диаграмма Венна состоит из частично перекрывающихся друг друга окружностей или других замкнутых кривых, демонстрирующих взаимосвязь и группировку множеств данных, которые могут быть общими для различных выборок или не пересекаться. Диаграммы Венна впервые были представлены Джоном Венном, преподавателем гуманитарных наук в Caius College Кембриджского университета, в статье под названием «О схематическом и механическом представлении суждений и рассуждений», опубликованной в 1880 г. [2]. Примечательно, что если проследить историю использования методов визуализации в области формальной логики, можно обнаружить, что диаграммы, ассоциируемые с именем Венна, по всей видимости, впервые были созданы существенно раньше [3]. Чтобы больше узнать об истории диаграмм Венна, читатели могут обратиться к другим публикациям [4].

По мере накопления знаний о сложности физиологических и молекулярных процессов все большее значение приобретает анализ мультиомики и больших данных (Big Data). Возрастает потребность разработки новых методов одновременной оценки большого числа наборов данных [5], в связи с чем даже классические методы визуализации, такие как диаграммы Венна, претерпели дальнейшее развитие. В настоящее время диаграммы Венна широко используют в клинических исследованиях, в частности для изучения микробиома [6] или обструктивных заболеваний легких [7], другой сферой их применения являются генетические исследования [8]. Графическое отображение взаимодействий легко воспринимается, поскольку отражает логические взаимосвязи и пересечения множеств.

В настоящее время для работы с четырьмя независимыми наборами данных доступны многочисленные онлайн-программы для построения диаграмм Венна, например генератор диаграмм Венна Pangloss<sup>1</sup> или Venny<sup>2</sup>. Эти программы весьма удобны

<sup>1</sup> <http://www.pangloss.com/seidel/Protocols/venn.cgi>

<sup>2</sup> <https://bioinfogp.cnb.csic.es/tools/venny/>

в применении, однако создают лишь изображения без значимой сопутствующей информации. BioVenn [9] является примером программ, позволяющих создавать диаграммы пропорциональной площади и проводить их анализ. Программы GeneVenn<sup>3</sup> и VennMaster<sup>4</sup> обладают дополнительной функцией привязки генов в каждой группе к ассоциированной информации в базах данных NCBI Entrez Nucleotide и Gene Ontology, однако в них можно проводить анализ лишь двух или трех панелей генов [8, 10].

Для баз данных микрочипов можно использовать специально разработанные для этих целей программы GeneSpring и SilicoCyte, в которых возможно и построение диаграмм Венна. Визуализация пересечений может создавать проблемы при работе с более чем тремя выборками из-за возможных нарушений симметрии изображений. Сам Венн в подобных случаях использовал дополнительные эллипсы, которые пересекались с основными окружностями [2]. Впоследствии A.W.F. Edwards разработал элегантный

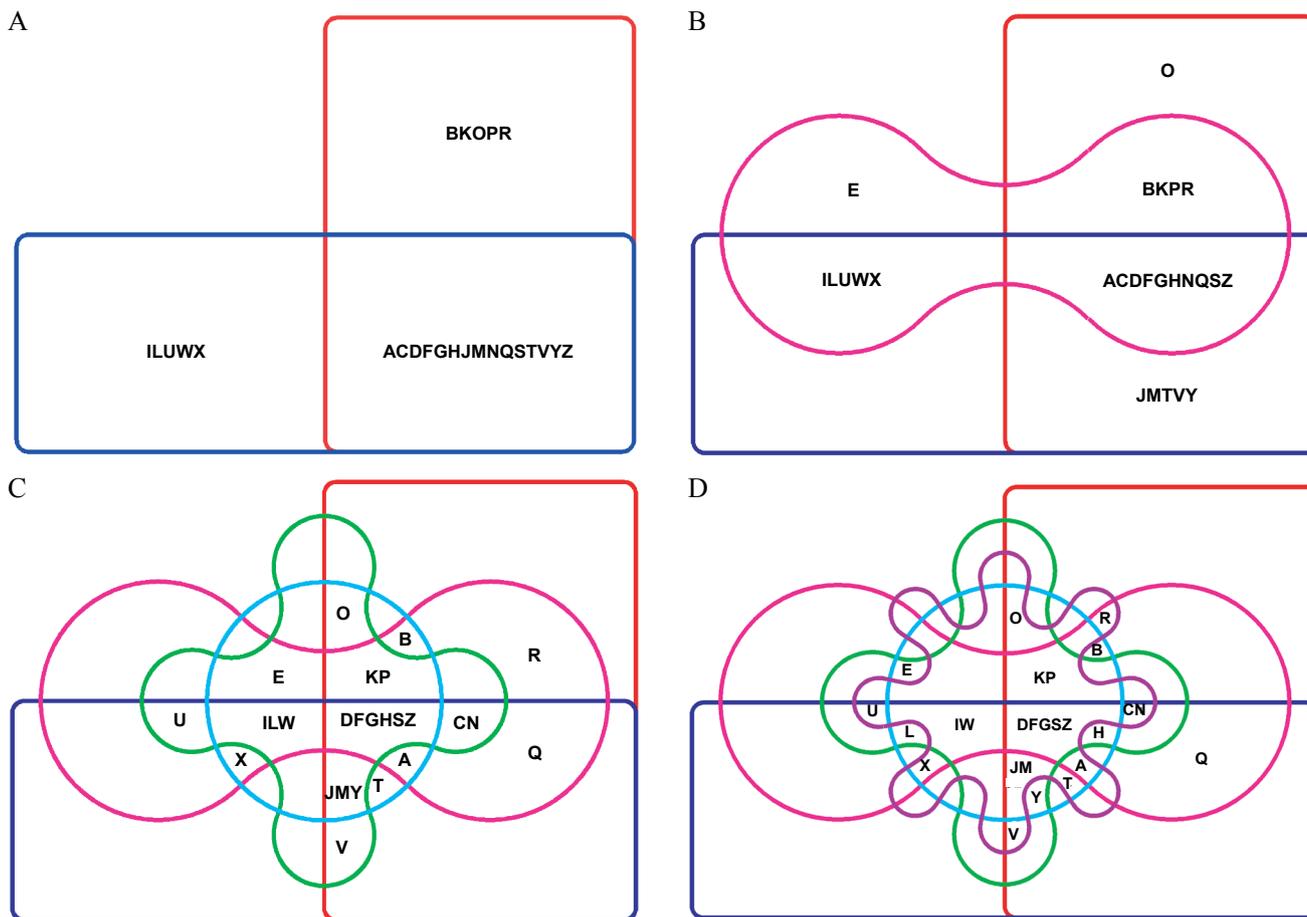
метод изображения диаграмм, включающих большее число множеств, с использованием необычных типов симметрии [11]. Далее представлен пример создания диаграмм Эдвардса — Венна, который можно с легкостью провести в свободно доступной программе VENNTURE<sup>5</sup> [5].

Используя программу VENNTURE, создадим диаграмму Эдвардса — Венна для абстрактного примера. Давайте представим, что в городе есть шесть аптек, которые могут продавать 26 наименований препаратов, и их можно обозначить буквами английского алфавита. Однако в ассортименте каждой аптеки может быть только 20 различных препаратов в связи с ограниченным местом хранения на складе. Предположим, что препараты были случайным образом распределены между аптеками. Используя диаграммы Венна, можно проанализировать, как именно они были распределены. Этот процесс изображен на рисунке 1 для 2 (A), 3 (B), 5 (C) и 6 (D) наборов данных, представляющих ассортимент препаратов в соответствующих аптеках. На рисунке 1A видно,

<sup>3</sup> <http://genevenn.sourceforge.net/>

<sup>4</sup> <https://sysbio.uni-ulm.de/?Software:VennMaster>

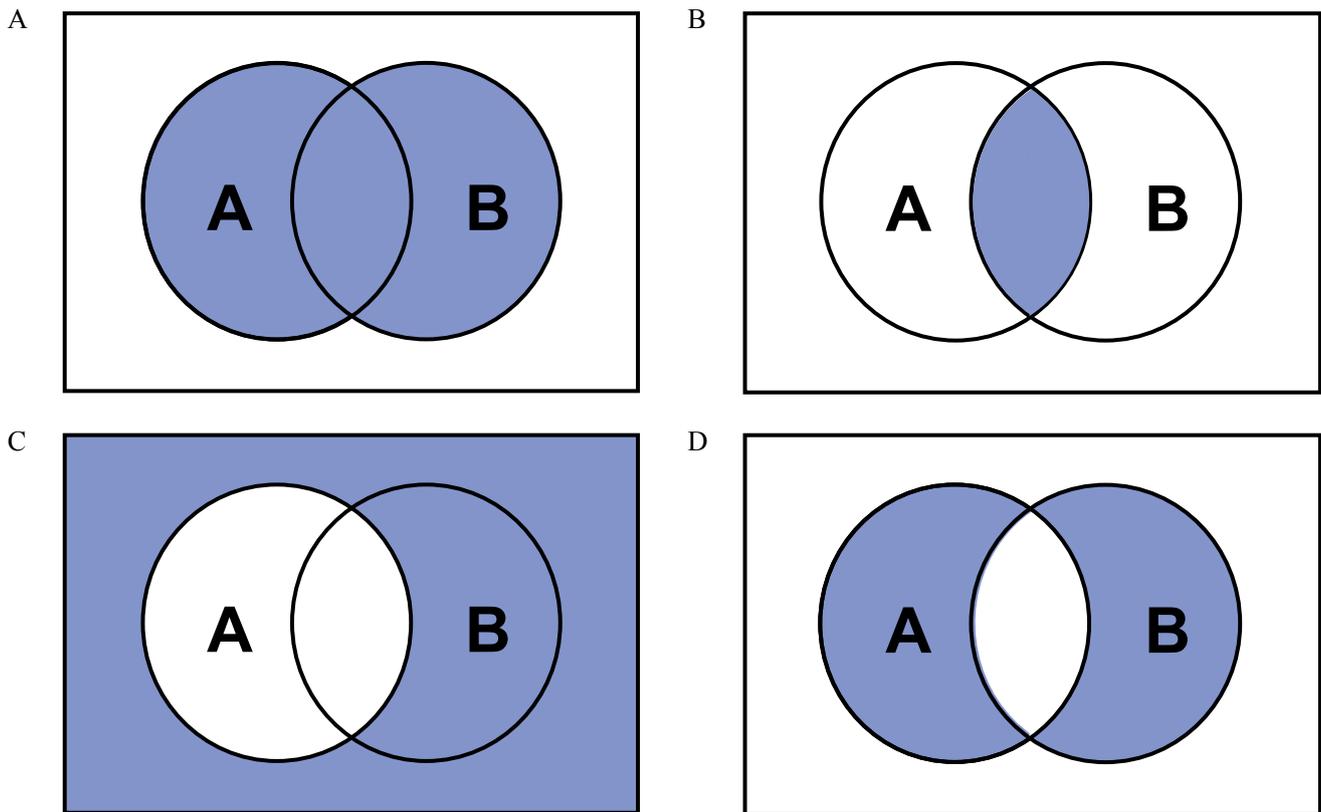
<sup>5</sup> <https://www.nia.nih.gov/research/resource/vennture>



**РИС. 1.** Диаграммы Эдвардса — Венна, симметрично изображающие пересечения 2 (A), 3 (B), 5 (C) и 6 (D) наборов данных.  
**FIG. 1.** Edwards—Venn diagram representing overlaps of 2 (A), 3 (B), 5 (C) and 6 (D) datasets in a symmetric way.

Примечание. Символы, которые являются общими для соответствующих наборов данных, изображены в области их пересечения.

Note. The symbols, which are common in the corresponding sets, are displayed at the intersections between these datasets.



**РИС. 2.** Диаграммы Венна, на которых синим цветом обозначено: объединение A и B, представленных двумя окружностями (A); пересечение A и B (B); событие, противоположное A, обозначаемое  $\bar{A}$  (C); симметричная разность A и B (D).  
**FIG. 2.** Venn diagram are represented in blue: a union of events A and B, shown by two circles (A); an intersection of events A and B (B); complement to event A denoted  $\bar{A}$  (C); and a symmetric difference of A and B (D).

что в обеих аптеках есть в наличии 15 препаратов, поскольку они попадают в пересечение наборов, обозначенных красным (первая аптека) и синим (вторая аптека) цветами. А пять препаратов есть в наличии только в одной из них. Предложенный метод визуализации позволяет отобразить до 6 наборов данных. Например, из рисунка 1B понятно, что препарат «О» можно найти только в первой аптеке, в то время как 10 препаратов продаются во всех трех исследуемых аптеках. Также становится понятно, что шесть препаратов можно найти в пяти аптеках (рис. 1C), а во всех шести аптеках — только пять препаратов (рис. 1D), поскольку в пересечение всех шести множеств попали только буквы «DFGSZ».

### РУКОВОДСТВО: ТЕОРИЯ МНОЖЕСТВ И ДИАГРАММЫ ВЕННА

Диаграммы Венна легче понять, если рассмотреть их связь с теорией вероятностей и теорией множеств. Вероятность — это способ выразить представление о каком-либо событии, которое произошло или может произойти. Любой процесс наблюдения событий в реальном мире можно назвать «экспериментом». Результаты эксперимента — это его исходы. В целом концепция вероятности нужна для формализации

принципов получения данных в наблюдениях. Любая модель, описанная в терминах вероятности («идеальный мир» модели), может быть изучена с помощью математических свойств, позволяющих понять закономерности и предсказать некоторые события, которые могут произойти в будущем или в других условиях (например, в клинических исследованиях). Статистика, напротив, изучает частоту уже произошедших в реальных условиях событий (например, ретроспективные данные или клинические исследования), чтобы оценить вероятность наблюдаемых событий и создать или протестировать подходящую модель вероятности.

Чтобы дать математическое определение вероятности, необходимо обозначить множество  $S$ , включающее в себя все возможные результаты эксперимента. Элемент «s» множества  $S$  называют элементарным событием. Пространство элементарных событий  $S$  называют:

- *дискретным*, если оно состоит из конечного числа элементарных событий;
- *исчислимым*, если все его элементы можно сопоставить с последовательностью положительных целых чисел (1, 2, 3, 4, ...);
- *непрерывным*, если все элементарные события составляют континуум;

• *пустым* или *нулевым* множеством, если множество не содержит элементов; подобные множества обозначают символом  $\emptyset$ .

Над множествами можно выполнять ряд операций, результаты которых можно отобразить с помощью диаграмм Венна (рис. 2).

### Объединение (рис. 2А)

События  $A$  и  $B$  в множестве  $S$  можно *объединить*, получив в результате более крупное событие. Объединение  $A \cup B$  возникает, когда случается одно событие или оба:

$$A \cup B = \{\text{все исходы в } A \text{ или в } B \text{ или в } A \text{ и } B \text{ вместе}\}$$

Если мы наблюдаем некое число событий  $E_i$ , равное  $n$  (где  $i$  — число от 1 до  $n$ ), их объединение обозначают:

$$\bigcup_{i=1}^n E_i = E_1 \cup E_2 \dots \cup E_n = \\ = \{\text{по крайней мере одно из } E_i \text{ произошло}\}$$

Объединение обладает следующими *свойствами*:

1. Объединение события  $A$  с пустым множеством равно событию  $A$  (закон тождества):

$$A \cup \emptyset = A$$

2. Объединение события  $A$  с самой собой равно событию  $A$  (закон идемпотентности/одинаковости):

$$A \cup A = A$$

3. Объединение события  $A$  с пространством элементарных событий равно пространству элементарных событий, или, иными словами, пространство элементарных событий доминирует над отдельным событием (закон доминирования):

$$A \cup S = S$$

4. Объединение событий  $A$  и  $B$  равно объединению событий  $B$  и  $A$  (коммутативный/переместительный закон):

$$A \cup B = B \cup A$$

Примеры:

1.1. Результаты подбрасывания монеты можно обозначить как  $A$ , если выпадает орел ( $A = \{O\}$ ), и  $B$ , если выпадает решка ( $B = \{P\}$ ). Объединение событий  $A$  и  $B$  будет включать в себя два элемента — выпадение и орлов, и решек ( $A \cup B = \{O, P\}$ ) и фактически равно всему пространству элементарных событий для данной модели.

1.2. Если обозначить результаты броска игральной кости при выпадении 1, 3 и 4 как событие  $E$  ( $E = \{1,3,4\}$ ) а при выпадении 1, 2 и 3 как событие  $F$  ( $F = \{1,2,3\}$ ), объединение событий  $E$  и  $F$  будет равно 1,2,3,4 ( $E \cup F = \{1,2,3,4\}$ ). Обратите внимание, что результаты 1 и 3 присутствуют в обоих событиях, но при объединении учитываются лишь однократно.

1.3. При игре в рулетку объединение всех нечетных результатов будет равно  $S = \{0\} \cup \{1,3, \dots, 33\}$ .

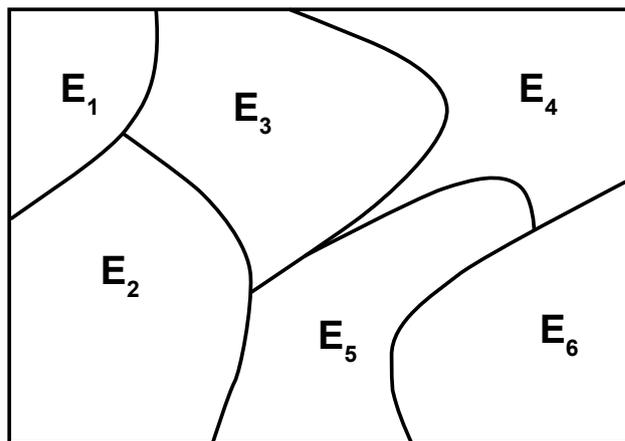


РИС. 3. События  $E_1$ – $E_6$  образуют разбиение пространства элементарных событий.

FIG. 3. Events  $E_1$ – $E_6$  form the partition of the sample space.

### Пересечение (рис. 2В)

Пересечение событий  $A \cap B$  позволяет создать новое событие, которое объединяет элементы, входящие одновременно и в  $A$ , и в  $B$  и возникающее, когда одновременно происходят события  $A$  и  $B$ .

$$A \cap B = \{\text{все исходы и в } A \text{ и в } B\}$$

Пересечение некоторого числа событий  $E_i$ , равного  $n$  (где  $i$  — число от 1 до  $n$ ) обозначают:

$$\bigcap_{i=1}^n E_i = E_1 \cap E_2 \dots \cap E_n = \\ = \{\text{все события } E_i \text{ произошли}\}$$

### Комплементарное (противоположное) событие и симметричная разность (рис. 2С и D)

Событие, противоположное  $A$ , или событие «не  $A$ » (обозначается  $\bar{A}$ ) включает в себя все исходы, которые не входят в  $A$  (рис. 2С). Развитием идеи комплементарного события является симметричная разность, которая для множеств  $A$  и  $B$  представляет все события в  $A$ , не входящие в  $B$ , и все события в  $B$ , не входящие в  $A$  (рис. 2D):

$$A \Delta B = (A \setminus B) \cup (B \setminus A)$$

### РУКОВОДСТВО: РАЗБИЕНИЕ ПРОСТРАНСТВА ЭЛЕМЕНТАРНЫХ СОБЫТИЙ

Используя две базовые операции, объединение и пересечение, можно выполнить разбиение пространства элементарных событий. Если пересечение событий  $A$  и  $B$  равно нулю ( $A \cap B = \emptyset$ ), эти события являются взаимоисключающими (несовместными). Взаимоисключающие события не могут происходить одновременно, или, другими словами, у этих множеств нет общих элементов. События называют исчерпывающими, если они включают в себя все

**Таблица 1. Эквиваленты операций над множествами в теории вероятностей**  
**Table 1. Summary of set operations and probability equivalents**

Выражение	Теория множеств	Теория вероятностей
S	Все исходы	Пространство элементарных событий
s	Элемент в S	Элементарное событие, исход
A	Подмножество S	Событие, при котором наблюдают исход A
$\bar{A}, A^c$	Событие, противоположное A	Событие, при котором исход A не происходит
$A \cap B$	Пересечение	И A, и B
$A \cup B$	Объединение	Или A, или B, или оба
$A \setminus B$	Разность	A, но не B
$A \Delta B$	Симметричная разность	Или A, или B, но не оба
$A \in B$	Инклюзия	Если A, то и B
$\emptyset$	Пустое множество	Невозможное событие
S	Все множество	Событие, которое произойдет наверняка

возможные исходы (подобно событиям A и B в примере 1.1 с монетой). Если события  $E_1, E_2, \dots, E_n$  взаимоисключающие и исчерпывающие, они формируют разбиение пространства элементарных событий (рис. 3).

Теория множеств позволяет интерпретировать различные ситуации, возникающие в реальном мире. A некоторые операции позволяют проследить взаимосвязи между терминологией теории вероятностей и теории множеств (табл. 1).

**РУКОВОДСТВО: ТЕОРИЯ ВЕРОЯТНОСТЕЙ**

Вероятность — мера количественной оценки возможности того, что событие произойдет. Возможны несколько интерпретаций вероятности. Первый подход — частотная оценка, которую получают путем вычисления доли случаев, в которых произошло некое событие, в долгосрочных сериях наблюдения при повторных экспериментах. Однако классическая интерпретация, предложенная российским математиком А.Н. Колмогоровым, определяет вероятность как величину, в равной степени распределенную между всеми возможными исходами. Таким образом, вероятность события — это доля от общего числа возможных исходов, в которых наступает исследуемое событие. Например, бросок монеты может привести лишь к двум возможным исходам: выпадению орла или решки, при этом вероятность каждого из этих событий равна 0,5.

Частотный подход к оценке вероятности (частотная вероятность) может быть проиллюстрирован следующим примером.

Предположим, что мы повторяем эксперимент N раз в абсолютно идентичных условиях. Событие A может произойти при каждом повторении эксперимента. Таким образом, если среди N наблюдений исход A наблюдали N(A) раз, то отношение N(A)/N стремится к вероятности этого события, поскольку N стремится к бесконечности:

$$\frac{N(A)}{N} \rightarrow p, \text{ поскольку } N \rightarrow \infty$$

В представленном примере p — постоянная величина, отражающая вероятность наступления события A в каждом отдельно взятом эксперименте — P(A). Эта величина обладает следующими свойствами.

1. Если A равно нулю, то событие никогда не происходит и его вероятность равна нулю:

$$\text{если } A = \emptyset, \text{ то } N(\emptyset) = 0, \text{ и } \frac{N(\emptyset)}{N} = 0$$

2. Если A включает в себя все возможные события, то исход наступает в каждом эксперименте, а его вероятность равна 1:

$$\text{если } A = S, \text{ то } N(S) = 1, \text{ и } \frac{N(S)}{N} = 1$$

3. Если A и B — несвязанные события, то частота их объединения равна сумме частот наблюдений события A — N(A), и события B — N(B), а вероятность наблюдения их объединения будет равна сумме вероятности наступления каждого события по отдельности:

$$\frac{N(A \cup B)}{N} = \frac{N(A) + N(B)}{N}$$

Однако этот же эксперимент можно рассмотреть и с позиции **классического** подхода. В таком случае эксперимент следует рассматривать как пространство элементарных событий S. Тогда для каждого события A, входящего в множество S, существует функция вероятности P(A), которая определяет вес A. Функция вероятности P(A) обладает следующими свойствами.

1. Ее величина находится в интервале от 0 до 1:

$$0 \leq P(A) \leq 1.$$

2. Вероятность совокупности всех возможных элементарных событий равна 1:

$$P(S) = 1.$$

3. Если события A и B не пересекаются (их пересечение равно 0), то вероятность их объединения равна сумме вероятностей наступления каждого события по отдельности:

$$\text{если } A \cap B = \emptyset, \text{ то } P(A \cup B) = P(A) + P(B)$$

Используя эти аксиомы, мы можем описать функцию вероятности. Для каждой серии

взаимоисключающих исходов  $A_1, A_2, \dots$  вероятность объединения будет равна сумме вероятностей каждого исхода по отдельности:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Следует помнить, что вероятности различных событий могут суммироваться. То есть, чтобы измерить вероятность множества событий, мы можем разделить его на конечное число отдельных событий, измерить вероятность каждого по отдельности и суммировать результаты.

Если событие  $A$  случается во всех наблюдениях, его вероятность равна 1 (100%):

$$P(A) = 1, \text{ если } A \text{ случается во всех наблюдениях.}$$

Если событие  $A$  невозможно (не случается никогда), его вероятность равна нулю:

$$P(A) = 0, \text{ если } A \text{ невозможно.}$$

Используя этот подход, рассмотрим несколько примеров.

### Вероятность противоположного

Рассмотрим событие  $A$ , представленное белым кругом на диаграмме Венна (рис. 2С). Давайте попробуем оценить вероятность противоположного события —  $\bar{A}$ , которое означает, что событие  $A$  не наступит — оно обозначено синим на диаграмме Венна. Используя ранее обсуждавшиеся утверждения, мы можем сделать следующие заключения.

Вероятность наступления всех возможных исходов равна 1:

$$P(S) = 1P(S) = 1. \quad (I)$$

Событие  $A$  и противоположное ему являются взаимоисключающими, поэтому их общая вероятность равна сумме вероятностей каждого из этих событий:

$$A \cap \bar{A} = 0 \text{ и } P(A \cup \bar{A}) = P(A) + P(\bar{A}) \quad (II)$$

События  $A$  и  $\bar{A}$  являются взаимоисключающими и исчерпывающими, поэтому их объединение включает в себя все пространство элементарных событий  $S$ :

$$(A \cup \bar{A}) = S(A \cup \bar{A}) = S \quad (III)$$

Обобщив заключения I–III, можно сделать вывод о том, что поскольку вероятность  $S$  равна вероятности объединения событий  $A$  и  $\bar{A}$ , которая, в свою очередь, равна сумме вероятностей каждого из этих событий, то сумма этих вероятностей равна 1:

$$1 = P(S) = P(A \cup \bar{A}) = P(A) + P(\bar{A})$$

Таким образом, в таких условиях вероятность развития события может быть вычислена через вероятность противоположного исхода:

$$P(\bar{A}) = 1 - P(A)$$

### Вероятность объединения

Рассмотрим два возможных события,  $A$  и  $B$  в множестве  $S$  (обозначены синим на диаграмме Венна, рис. 2А). Давайте оценим вероятность наступления какого-либо из двух исходов,  $A$  или  $B$ , которая, очевидно, равна вероятности их объединения.

Существенной проблемой является пересечение  $A$  и  $B$ , потому что, просто сложив вероятности  $A$  и  $B$ , мы учтем исходы в области пересечения дважды, что неверно. Чтобы оценить вероятность пересечения точно, следует принять во внимание следующее.

Объединение  $A$  и  $B$  равно объединению  $A$  с областью пересечения противоположного события  $\bar{A}$  и события  $B$ :

$$A \cup B = A \cup (\bar{A} \cap B)$$

В свою очередь, вероятность объединения  $A$  и  $B$  равна сумме вероятностей этих двух множеств:

$$P(A \cup B) = P(A) + P(\bar{A} \cap B) \quad (I)$$

С другой стороны,  $B$  можно представить в виде объединения области пересечения  $A$  и  $B$  с областью пересечения пространства вне события  $A$  ( $\bar{A}$ ) и  $B$ :

$$B = (A \cap B) \cup (\bar{A} \cap B)$$

$$P(B) = P(A \cap B) + P(\bar{A} \cap B) \quad (II)$$

С помощью выражений I и II мы можем переписать уравнения и заменить их члены, чтобы представить вероятность объединения событий как сумму их вероятностей за вычетом вероятности их пересечения:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

### Многokrатное подбрасывание монеты

Монету, вероятность выпадения как орла, так и решки у которой равна 0,5, подбрасывают несколько раз. Используя теорию вероятностей, можно доказать, что наверняка (с вероятностью, равной 1) рано или поздно орел выпадет.

Сначала можно оценить вероятность противоположного события — орел никогда не выпадет. Она равна пределу  $n$  ( $n$  стремится к бесконечности) вероятности того, что за первые  $n$  бросков орел не выпадет. Вероятность того, что будет выпадать только решка, равна  $1/2^n$ . Поскольку  $n$  стремится к бесконечности, значение этого уравнения стремится к нулю:

$$P(\text{орел никогда не выпадет}) = \lim_{n \rightarrow \infty} P(\text{орел не выпадет в первые } n \text{ бросков}) = \lim_{n \rightarrow \infty} (2^{-n}) = 0$$

Вероятность выпадения орла равна единице минус вероятность противоположного события (орел не выпадет), которая стремится к нулю. Таким образом, искомая вероятность равна:

$$P(\text{орел выпадет}) = 1 - P(\text{орел не выпадет}) = 1 - 0 = 1$$

Вероятность события, равная единице, означает, что орел рано или поздно выпадет.

**Таблица 2. Пример анализа продолжительности жизни клеток**

*Table 2. Lifetime of cells analysis example*

Продолжительность жизни, ч	Число клеток	Доля
<1000	45	0,225
1000–1500	80	0,400
>1500	75	0,375

### Продолжительность жизни клеток

Последний пример демонстрирует применение теории вероятностей в биомедицинских исследованиях. Эксперимент состоит в измерении продолжительности жизни 200 клеток. Результаты представлены в таблице 2.

Наша задача состоит в оценке вероятности того, что продолжительность жизни клетки составит  $\leq 1500$  ч. Сначала необходимо обозначить события: К — короткое, для исхода  $< 1000$  часов; С — среднее, для исходов в промежутке от 1000 до 1500 часов; Д — длинное, для исходов  $> 1500$  часов.

Чтобы оценить вероятность того, что продолжительность жизни составит  $\leq 1500$  часов, необходимо оценить вероятность объединения событий К и С, которая равна сумме их вероятностей, поскольку их пересечение равно нулю (взаимоисключающие события):

$$P(h \leq 1500) = P(K \cup C) = P(K) + P(C) = 0,225 + 0,4 = 0,625$$

### ВКЛАД АВТОРОВ

Н.М. Буланов, А.А. Заикин, О.Б. Блюсс, Д.Б. Мунблит, Т.В. Назаренко и М.Ю. Надинская участвовали в написании текста рукописи. О.Б. Блюсс, Д.Б. Мунблит и Т.В. Назаренко выполняли поиск и анализ литературы по теме обзора. А.А. Заикин и Д.В. Бутнару разработали общую концепцию статьи и осуществляли руководство ее написанием. Все авторы участвовали в обсуждении и редактировании работы. Все авторы утвердили окончательную версию публикации.

### ЛИТЕРАТУРА / REFERENCES

- 1 Kirkwood B.R., Sterne J.A.C. Essential Medical Statistics (2nd ed.), Wiley-Blackwell, 2003. 512 p.
- 2 Venn J. On the diagrammatic and mechanical representation of propositions and reasonings. The London, Edinburgh, and Dublin philosophical magazine and journal of science, 1880; 10(59): 1–18. <https://doi.org/10.1080/14786448008626877>
- 3 Bultena B., Ruskey F. Venn diagrams with few vertices. Electronic Journal of Combinatorics. 1998; 5: R44: 1–21. <https://doi.org/10.37236/1382>
- 4 Baron M. E. A note on the historical development of logic diagrams: Leibniz, Euler and Venn. The Mathematical Gazette. 1969; 53(384): 113–125. <https://doi.org/10.2307/3614533>
- 5 Martin B., Chadwick W., Yi T., et al. VENNTURE — a novel Venn diagram investigational tool for multiple pharmacological dataset analysis. PLoS One. 2012; 7(5): e36911. <https://doi.org/10.1371/journal.pone.0036911> Erratum in: PLoS One. 2012; 7(5): <https://doi.org/10.1371/annotation/27f1021c-b6f2-4b90-98bc-fcacd2679185> PMID: 22606307
- 6 Shade A., Handelsman J. Beyond the Venn diagram: the hunt for a core microbiome. Environ Microbiol. 2012 Jan; 14(1): 4–12. <https://doi.org/10.1111/j.1462-2920.2011.02585.x> PMID: 22004523.
- 7 Soriano J.B., Davis K.J., Coleman B., et al. The proportional Venn diagram of obstructive lung disease: two approximations from the United States and the United Kingdom. Chest. 2003 Aug; 124(2):474–81. <https://doi.org/10.1378/chest.124.2.474> PMID: 12907531
- 8 Pirooznia M., Nagarajan V., Deng Y. GeneVenn — A web application for comparing gene lists using Venn diagrams. Bioinformatics. 2007 Apr 10; 1(10): 420–2. <https://doi.org/10.6026/97320630001420> PMID: 17597932
- 9 Hulsen T., de Vlieg J., Alkema W. BioVenn — a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. BMC Genomics. 2008 Oct 16; 9: 488. <https://doi.org/10.1186/1471-2164-9-488> PMID: 18925949
- 10 Kestler H.A., Müller A., Kraus J.M., et al. VennMaster: area-proportional Euler diagrams for functional GO analysis of microarrays. BMC Bioinformatics. 2008 Jan 29; 9: 67. <https://doi.org/10.1186/1471-2105-9-67> PMID: 18230172
- 11 Edwards A. Venn diagrams for many sets. New Scientist, 1989; 121(1646): 51–6.

Обратите внимание, что события К, С и Д не только взаимоисключающие, но и исчерпывающие, а потому образуют разбиение пространства элементарных событий.

### ЗАКЛЮЧЕНИЕ

В первой вводной статье мы представили краткий обзор области применения диаграмм Венна в современных клинических исследованиях. Следует отметить, что даже этот метод визуализации, который является классическим, продолжает активно развиваться. Мы рассмотрели базовые положения теории вероятностей, ее взаимосвязь с теорией множеств, а также применение диаграмм Венна для иллюстрации некоторых концепций. Несмотря на то что материал может показаться довольно абстрактным, он имеет критическое значение для понимания принципов статистического анализа, а диаграммы Венна нередко используют для иллюстрации в научных статьях.

### AUTHOR CONTRIBUTIONS

Nikolay M. Bulanov, Alexey A. Zaikin, Oleg B. Blyuss, Daniil B. Munblit, Tatiana V. Nazarenko and Maria Yu. Nadinskaia, participated in writing the text of the manuscript. Oleg B. Blyuss, Daniil B. Munblit and Tatiana V. Nazarenko searched and analyzed the literature on the review topic. Alexey A. Zaikin and Denis V. Butnaru developed the general concept of the article and supervised its writing. All authors participated in the discussion and editing of the work. All authors approved the final version of the publication.

## ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

**Буланов Николай Михайлович** , канд. мед. наук, ассистент кафедры внутренних, профессиональных болезней и ревматологии ФГАОУ ВО «Первый МГМУ им. И.М. Сеченова» Минздрава России (Сеченовский Университет).  
ORCID: <https://orcid.org/0000-0002-3989-2590>

**Блюсс Олег Борисович**, канд. физ.-мат. наук, доцент кафедры педиатрии и детских инфекционных болезней ФГАОУ ВО «Первый МГМУ им. И.М. Сеченова» Минздрава России (Сеченовский Университет).  
Старший преподаватель, Школа физики, астрономии и математики, Университет Хартфордшира.  
ORCID: <https://orcid.org/0000-0002-0194-6389>

**Мунблит Даниил Борисович**, PhD, профессор кафедры педиатрии и детских инфекционных болезней ФГАОУ ВО «Первый МГМУ им. И.М. Сеченова» Минздрава России (Сеченовский Университет).  
Почетный старший преподаватель, Секция воспаления, регенерации и развития, Национальный институт сердца и легких, Медицинский факультет, Имперский колледж Лондона.  
ORCID: <https://orcid.org/0000-0001-9652-6856>

**Назаренко Татьяна Валерьевна**, научный сотрудник, Институт женского здоровья (Женская онкология), Университетский колледж Лондона.  
ORCID: <https://orcid.org/0000-0002-4245-7346>

**Бутнару Денис Викторович**, канд. мед. наук, проректор по научной работе ФГАОУ ВО «Первый МГМУ им. И.М. Сеченова» (Сеченовский Университет).  
ORCID: <https://orcid.org/0000-0003-2173-0566>

**Надинская Мария Юрьевна**, канд. мед. наук, доцент кафедры пропедевтики внутренних болезней, гастроэнтерологии и гепатологии ФГАОУ ВО «Первый МГМУ им. И.М. Сеченова» Минздрава России (Сеченовский Университет).  
ORCID: <http://orcid.org/0000-0002-1210-2528>

**Заикин Алексей Анатольевич**, канд. физ.-мат. наук, заместитель директора Центра анализа сложных систем ФГАОУ ВО «Первый МГМУ им. И.М. Сеченова» Минздрава России (Сеченовский Университет).  
Профессор системной медицины, Институт женского здоровья и кафедра математики, Университетский колледж Лондона.  
ORCID: <https://orcid.org/0000-0001-7540-1130>

**Nikolay M. Bulanov** , Cand. of Sci. (Medicine), Assistant Professor, Department of Internal, Occupational Diseases and Rheumatology, Sechenov First Moscow State Medical University (Sechenov University).  
ORCID: <https://orcid.org/0000-0002-3989-2590>

**Oleg B. Blyuss**, Cand. of Sci. (Phys. and Math.), Associate Professor, Department of Paediatrics and Paediatric Infectious Diseases, Sechenov First Moscow State Medical University (Sechenov University).  
Senior Lecturer, School of Physics, Astronomy and Mathematics, University of Hertfordshire.  
ORCID: <https://orcid.org/0000-0002-0194-6389>

**Daniil B. Munblit**, PhD, Professor, Department of Paediatrics and Paediatric Infectious Diseases, Sechenov First Moscow State Medical University (Sechenov University).  
Honorary Senior Lecturer, Inflammation, Repair and Development Section, National Heart and Lung Institute, Faculty of Medicine, Imperial College London.  
ORCID: <https://orcid.org/0000-0001-9652-6856>

**Tatiana V. Nazarenko**, Research Fellow, Institute for Women's Health (Women's Cancer), University College London.  
ORCID: <https://orcid.org/0000-0002-4245-7346>

**Denis V. Butnaru**, Cand. of Sci. (Medicine), Vice-rector for Research, Sechenov First Moscow State Medical University (Sechenov University).  
ORCID: <https://orcid.org/0000-0003-2173-0566>

**Maria Yu. Nadinskaia**, Cand. of Sci. (Medicine), Associate Professor, Department of Internal Medicine Propaedeutics, Gastroenterology and Hepatology, Sechenov First Moscow State Medical University (Sechenov University).  
ORCID: <http://orcid.org/0000-0002-1210-2528>

**Alexey A. Zaikin**, Cand. of Sci. (Phys. and Math.), Deputy Director, Centre for Analysis of Complex Systems, Sechenov First Moscow State Medical University (Sechenov University).  
Professor of Systems Medicine, Institute for Women's Health and Department of Mathematics, University College London.  
ORCID: <https://orcid.org/0000-0001-7540-1130>

 Автор, ответственный за переписку / Corresponding author