



## 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](mailto: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

## 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

Statistics is the science of the collecting, summarizing, presenting and interpreting of data; using them to estimate the magnitude of associations and test hypotheses [1]. Today statistics has a leading role in medical and biological research. Therefore, it is crucial for all doctors and researchers to understand and be able to apply at least the basic concepts of data analysis to their studies. However, most healthcare providers in Russia have almost no training in mathematics and a very limited understanding of statistics resulting in the misinterpretation and uncritical acceptance of published medical data, as well as an inability to organize, perform and analyze their own research properly. For this reason, we decided to publish this introductory course on biomedical statistics in a series of articles.

This course was originally designed for Master and PhD students at University College London and was later adapted for students and postgraduate trainees at Sechenov University. The main idea of the course is to take a practical approach, and this is why the current and further articles will focus on the most important concepts, techniques and statistical tests required for the critical understanding of state-of-art clinical research and the preparation of manuscripts for submission to high-ranking medical journals.

We encourage the readers to learn to present data adequately, apply the basic statistical tests properly and interpret their results without help from a professional statistician. Nevertheless, we have also included in this course some advanced statistical methods such as basic machine-learning algorithms in order to acquaint the reader with their possible applications. Another aim of the course is to achieve efficient cooperation between doctors and statisticians in terms of creating rational clinical research study designs, formulating realistic inquiries, choosing optimal approaches and interpreting the results of the research. All of the discussed methods can be universally applied in any statistical software, both commercial (SPSS, Stata, SAS, MatLab, etc.) and freeware (R).

The material is designed to be understandable by the readers who are not mathematical specialists. However, some mathematical concepts and expressions will be introduced throughout the course in a simplified form since the authors believe it is the most adequate and accurate way to present the underlying ideas of statistical techniques. It was decided to present the course both in English and Russian to familiarize readers with the terms frequently used in scientific publications.

The first article will discuss the concepts of the set theory and probability theory with some examples illustrated by Venn diagrams.

## VENN DIAGRAMS

Venn diagrams (also called Set diagrams) are widely used in clinical and population health research reports to represent the logical relationships between two or more sets of data. A Venn Diagram involves overlapping circles or other closed curves, which present how things are organized, including their common and different sets. Venn diagrams were developed in 1880 by John Venn, lecturer in Moral Science at Caius College, Cambridge University in a paper entitled, "On the diagrammatic and mechanical representation of propositions and reasonings" [2]. It is noteworthy that the use of diagrams in formal logic is easily traced historically, and the diagrams normally associated with Venn, probably originated much earlier [3]. More information on the history of Venn diagrams can be found in other publications [4].

As our knowledge of the complexity of physiological and molecular functions expands, the study of multi-omics and Big Data become more and more important. There is an increasing need for novel methods to study multiple datasets [5], for even such a classical method of visualization as the Venn diagram has undergone further development. Venn diagrams are currently widely used in clinical research, for example, to study a core microbiome [6], or in obstructive lung diseases [7], and even genetic studies often include them. For example, one of the most effective ways of examining the overlap of resulting gene lists is a Venn diagram [8]. The graphic way to represent interactions can be read easily because it shows all the logical relations and overlaps between the sets.

Numerous online Venn programs are currently available to work with up to four independent datasets; with examples being Pangloss Venn diagram generator<sup>1</sup> or Venny<sup>2</sup>. These programs provide a convenient resource but only generate images without the relevant output information. BioVenn [9] is another available program allowing the generation of area-proportional Venn set analyses. Programs such as GeneVenn<sup>3</sup> and VennMaster<sup>4</sup> have the additional feature of linking genes within each group to related information in the NCBI Entrez Nucleotide database or the Gene Ontology database. However, these programs are still limited to the analysis of two or three gene lists [8, 10].

For microarray datasets GeneSpring and SilicoCytte can be used, both of which are specifically designed for microarray data with several visualizations, including Venn diagrams. For more than three data sets, the intersections can start to become problematic in terms of visualization as the representation of all overlapping

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

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

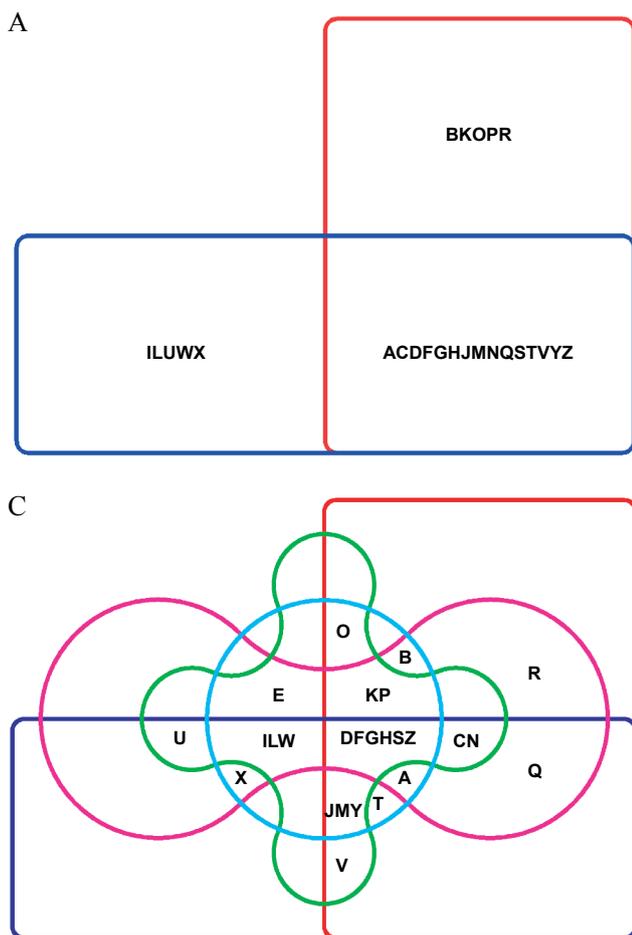
<sup>3</sup> <http://genevenn.sourceforge.net/>

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

regions and symmetry can be lost. Venn himself performed visualization using additional ellipses that intersect with the primary circles [2]. A.W.F. Edwards also developed an elegant method of displaying the diagram composed of higher number of sets that feature some interesting symmetries [11]. VENNTURE<sup>5</sup> [5] is a novel freely available Venn diagram-generating tool, and is a good example of an Edwards—Venn diagram, which can easily be performed with the program.

Using the VENNTURE program can allow us to generate the Edward—Venn diagram for an abstract example: Let us assume that there are 6 pharmacies in the city that can sell 26 types of drugs and let us designate these drugs with the letters of the English alphabet. However, each pharmacy can only accommodate 20 different types of drugs due to storage restrictions. If the drugs were randomly distributed between pharmacies, then one could use the Venn diagrams to analyse how the different drugs were distributed between the pharmacies. This is seen in the Fig. 1 for 2 (A),

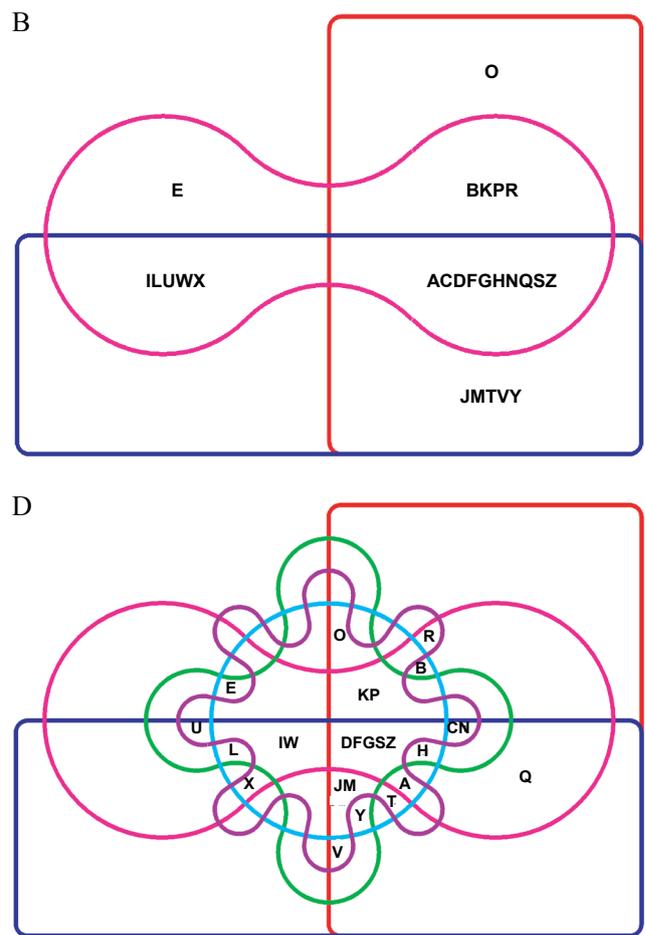
<sup>5</sup> <https://www.nia.nih.gov/research/resource/vennture>



3 (B), 5 (C) and 6 (D) datasets representing current list of letters-drugs in the corresponding pharmacies. From Fig. 1A we can see that 15 drugs are offered in both pharmacies because they are at the intersection of the datasets corresponding to the first (red) and second (blue) pharmacy. Five drugs can only be found in one of the two pharmacies. Suggested visualization method enables representation for up to 6 datasets. For example, Fig. 1B shows that drug “O” can only be found in the first pharmacy, whereas 10 drugs are being sold in all three pharmacies under analysis. One can also see that only 6 drugs can be found in 5 pharmacies (Fig. 1C) and 5 drugs are sold on all 6 pharmacies (Fig. 1D) because only the letters “DFGSZ” are at the intersection area of all 6 datasets.

**TUTORIAL: SET THEORY AND VENN DIAGRAMS**

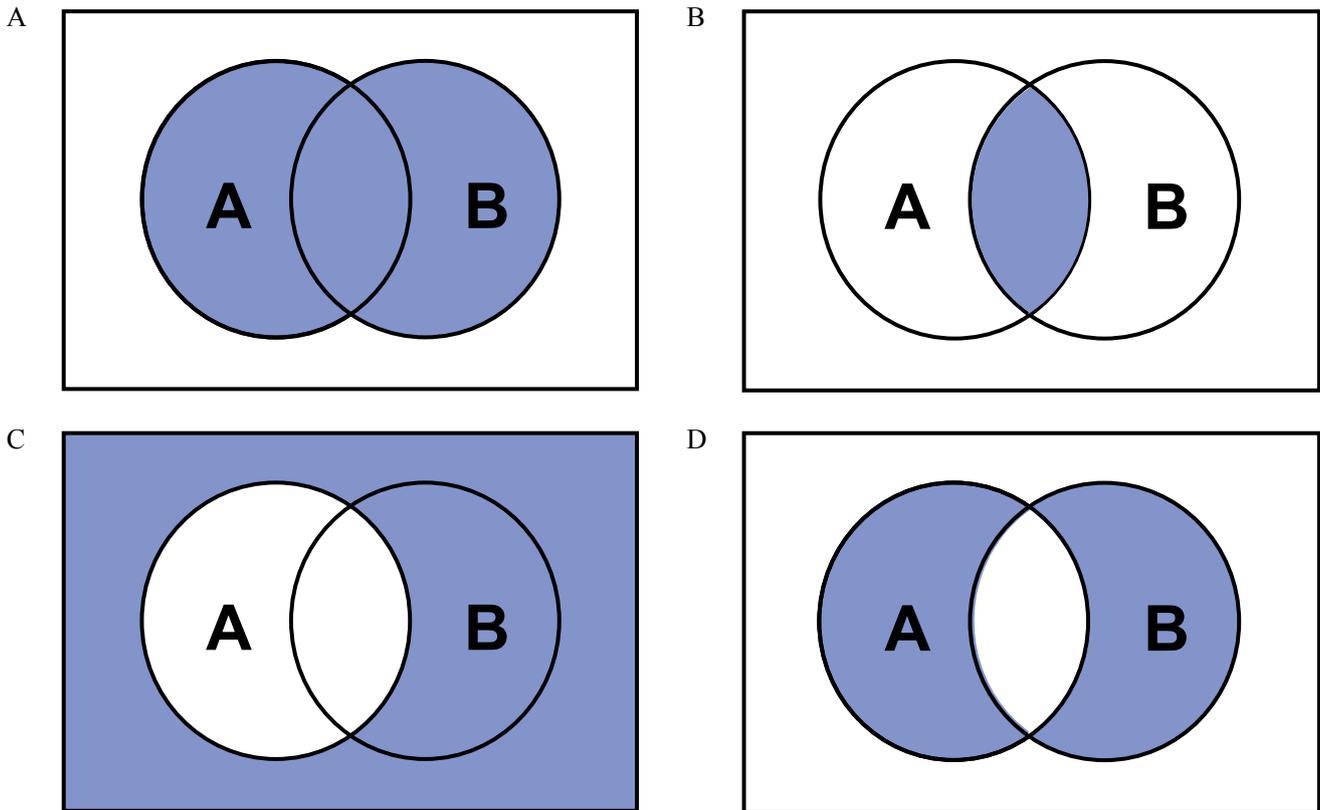
Venn diagrams can be best understood if we review the concept of probability and the link between the theory of sets and Venn diagrams. Probability is



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

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

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



**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).  
**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).

a way of expressing knowledge or belief that some event has occurred or will occur. Any process of observing the events in the real-world setting is referred to as “experiment”. The results of experiments are the outcomes. In general, probability is required to formalize the mechanisms that generate the data we get from the observations. Any model given in terms of probability (“ideal world” of the model) can be studied through the mathematical properties which can help understand the patterns and predict some events that are likely to happen in the future or under different settings (e.g., in a clinical trial). In contrast, statistics studies the frequency of past events that have happened in real world settings (e.g., historical data or a clinical trial) to assess the probability of observed events and create or test a suitable probability model.

To define probability mathematically we need to define set S consisting of all possible outcomes of the experiment. An element “s” of S is called a sample point. A sample space S is called:

- *discrete* if it consists of a finite number of sample points;

- *countable* if the elements can be placed in one-to-one correspondence with positive integers (1, 2, 3, 4, ...);
- *continuous* if the sample points constitute a continuum;
- *empty* or *null set* if it contains no elements; in that case this set is denoted by  $\emptyset$ .

Several operations can be performed with the sets of outcomes and the results can be graphically represented with the Venn diagrams (Fig. 2):

**Union of events (Fig. 2A)**

Given events A and B in set S, we can build a new larger event as their *union*. The union  $A \cup B$  occurs when at least one of these two events occurs:

$$A \cup B = \{\text{all outcomes in either A or in B or in both}\}$$

If we have a number of events  $E_i$  equal to n (where i is a number from 1 to n) their union is denoted as:

$$\bigcup_{i=1}^n E_i = E_1 \cup E_2 \dots \cup E_n = \\ = \{\text{at least one of the } E_i \text{ occurred}\}$$

The union has the following *properties*:

1. The union of event A with an empty set is event A (identity law):

$$A \cup \emptyset = A$$

2. The union of event A with itself is event A (idempotent law):

$$A \cup A = A$$

3. The union of event A with the sample space is the sample space, or in other words, the event is dominated by the sample space (domination law):

$$A \cup S = S$$

4. The union of A and B is exactly the same as the union of B and A (commutative law):

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

Examples:

1.1. If one tosses a coin the resulting events can be denoted as A corresponding to heads ( $A = \{H\}$ ) and B corresponding to tails ( $B = \{T\}$ ). The union of A and B will include two elements — both heads and tails ( $A \cup B = \{H, T\}$ ), corresponding to the whole sample space.

1.2. If one rolls a die and denote the results 1, 3 and 4 as event E ( $E = \{1,3,4\}$ ) and results 1, 2 and 3 as event F ( $F = \{1,2,3\}$ ), then the union of these two event E and F will be 1,2,3,4 ( $E \cup F = \{1,2,3,4\}$ ). Note that results 1 and 3 are presented in both sets but only included once.

1.3. If one is playing roulette, the union of all uneven results and zero is  $S = \{0\} \cup \{1,3, \dots, 33\}$

### Intersection of events (Fig. 2B)

Given events A and B, we can build a new event as their intersection  $A \cap B$ , which includes all the outcomes in both A and B and occurs when both A and B occur.

$$A \cap B = \{\text{all outcomes in both A and B}\}$$

Note that the intersection of multiple events  $E_i$  equal to n (where  $i$  is a number from 1 to n) is denoted as:

$$\bigcap_{i=1}^n E_i = E_1 \cap E_2 \dots \cap E_n = \{\text{all of the } E_i \text{ occurred}\}$$

### Complement and symmetric difference (Fig. 2C and 2D)

Complement of event A is the event “not A” ( $\bar{A}$ ), which includes all outcomes not present in the event A (Fig. 2C). An extension of the complement is the symmetric difference, which represents for sets A and B all outcomes in A without B and in B without A (Fig. 2D):

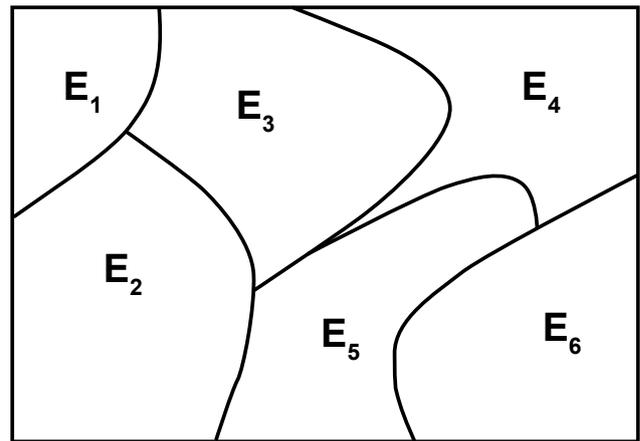


FIG. 3. Events  $E_1$ – $E_6$  form the partition of the sample space.

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

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

### TUTORIAL: PARTITION OF THE SAMPLE SPACE

Using two basic operations, union and intersection, the partition of the sample space can be defined. If the intersection of two events A and B is zero ( $A \cap B = \emptyset$ ), then these events are called mutually exclusive. Mutually exclusive events cannot happen together, or in other words, have no common elements. Events are called exhaustive if they include all possible outcomes (like the union of events A and B in example 1.1 with coins). If events  $E_1, E_2, \dots, E_n$  are mutually exclusive and exhaustive and are said to form a partition of sample space (Fig. 3).

We can use the set theory to describe the aspects of the real world in a manipulable notation. Some operations can help to reveal relations between set and probability terminology (Table 1).

### TUTORIAL: PROBABILITY THEORY

Probability is a measure of quantifying the likelihood that events will occur. There are several possible interpretations of probability. The first one is frequentist, which is based on long-run sequences in repeated trials or, in other words, on counting the proportion of occasions some event occurred. However, the classical interpretation, given by a Russian mathematician A.N. Kolmogorov, defines probability as a measure shared equally among all the possible outcomes. So the classical probability of an event is the fraction of the total number of possibilities in which the event occurs. For example, if one tosses a coin which can only result in two possible outcomes, heads or tails, the probability of each event will be 0.5.

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

Notation	Set terminology	Probability terminology
S	All outcomes	Sample space
s	Point in S	Elementary event, outcome
A	Subset of S	Event where some outcome in A occurs
$\bar{A}, A^c$	Complement of A	Event where no outcome in A occurs
$A \cap B$	Intersection	Both A and B
$A \cup B$	Union	Either A or B, or both
$A \setminus B$	Difference	A, but not B
$A \Delta B$	Symmetric difference	Either A or B, but not both
$A \in B$	Inclusion	If A, then B
$\emptyset$	Empty set	Impossible event
S	Whole space	Certain event

The frequentist approach can be illustrated with a following example:

We repeat an experiment N number of times under exactly the same conditions. Event A may occur or may not occur on each repetition. If N(A) is a number of times A occurs in these N observations, then we would expect that the ratio of N(A)/N would tend to probability (p) as N tends to infinity:

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

In that case p is the constant showing the probability of event A occurrence in any particular experiment — P(A). In that case this notion has the following properties.

1. If A is equal to zero, than the event never occurs and its probability is equal to 0:

$$\text{if } A = \emptyset, \text{ then } N(\emptyset) = 0, \text{ and } \frac{N(\emptyset)}{N} = 0$$

2. If A includes all sample space, then the event occurs in every experiment and its probability is equal to 1:

$$\text{if } A = S, \text{ then } N(S) = N, \text{ and } \frac{N(S)}{N} = 1$$

3. If A and B are disjoint events, then number of occasions we see their union will be equal to the sum of N(A) and N(B), so their total probability will be equal to the sum of probabilities:

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

To describe this experiment, we can also use the **classical** approach, where the experiment is considered as sample space S. For event A, which belongs to the sample space S, we assume there is a set probability

function P(A) that assigns a weight to A. Probability function P(A) satisfies the following properties:

1. It should be between zero and 1:

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

2. Probability of the whole sample space should be equal to 1:

$$P(S) = 1.$$

3. If two events A and B do not overlap (their intersection is equal to 0), then the probability of their union will be equal to the sum of probabilities of each event:

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

Using these axioms, we can produce the probability function. And for each sequence of events  $A_1, A_2, \dots$  that are mutually exclusive, the probability of the union will be equal to the sum of their individual probabilities:

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

One should keep in mind that probability is countably additive. It means that to measure a set we can decompose it into finitely many disjoint parts (events), measure each part separately and then add up the results. If some event A is certain (happens in all observations), then its probability is equal to 1 (100%):

$$P(A) = 1, \text{ if } A \text{ is certain}$$

If event A is impossible (never happens), then its probability is equal to zero:

$$P(A) = 0, \text{ if } A \text{ is impossible}$$

Using this approach, we can consider several examples.

### Probability of the complement

Consider event A, represented by a white circle on the Venn diagram (Fig. 2C). Let us assess the probability of the complement of this event —  $\bar{A}$ , which denotes that no outcome in A occurs — it is

given in blue on the Venn diagram. By using the previously discussed postulates we can conclude the following:

The probability of the total sample space is equal to 1:

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

Event A and its complement are mutually exclusive so their total probability is equal to the sum of probabilities of each event:

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

Events A and  $\bar{A}$  are mutually exclusive and exhaustive so their union is equal to the total sample space S:

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

Considering steps I–III we can conclude that since the probability of S is equal to the probability of the union of events A and  $\bar{A}$ , which in itself is equal to the sum of their probabilities, then the sum of probabilities of these events is equal to 1:

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

Therefore, in this setting, the probability of an event can be derived by the probability of its complement:

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

### Probability of the union

Consider two possible events A and B in the set S (shown in blue on the Venn diagram, Fig. 2A). Let us assess the probability of outcome in either A or B, which obviously is the probability of their union. The problem is that an intersection between A and B exists. That is why if we just add up the probabilities of A and B we would count this section twice, which is wrong. To measure the probability of a union properly we should take into account the following:

The union of A and B is equal to the union of A and the intersection between the area outside A (its complement,  $\bar{A}$ ) and event B:

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

And the probability of the union of A and B is equal to the sum of these probabilities:

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

On the other hand, the event B is equal to the union of intersection of A and B and intersection of area outside A ( $\bar{A}$ ) and B:

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

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

Using statements I and II, we can rearrange the equations and substitute their parts to reach the expression for the probability of the union between events, which is equal to the sum of probabilities of A and B minus probability of their intersection:

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

### Repeated coin tossing

A fair coin (probability of getting head or tail is equal to 0.5 each) is tossed repeatedly. Using the probability theory, we may show that with total certainty, i.e. with probability equal to 1, head turns up sooner or later.

First, we need to assess the probability of the complement — that no head will ever appear. It is equal to the limit of n (n tends to infinity) the probability that no head occurs in the first n tosses. The probability that we get only tails is equal to  $\frac{1}{2}^n$  times. As n tends to infinity, this equation tends to 0:

$$P(\text{no head ever}) = \lim_{n \rightarrow \infty} P(\text{no head for the first } n \text{ tosses}) = \lim_{n \rightarrow \infty} (2^{-n}) = 0$$

The probability that head turns up is equal to one minus its complement (no head appears), which tends to zero. Therefore, the required probability is equal to:

$$P(\text{head turns up}) = 1 - P(\text{no head ever}) = 1 - 0 = 1$$

Probability equal to 1 means that head will turn up with total certainty sooner or later.

### Lifetime of cells

The last example demonstrates a simple implementation of probability theory in biomedical research.

The experiment consisted of measuring the lifetime of 200 cells (table 2). The results are shown in table 2.

Our task is to compute the probability that a cell lifetime is less than or equal to 1500 hours. First, we need to define the events: Sh — short, for event lasting less than 1000 hours; M — medium, for event lasting between 1000 and 1500 hours; L — long, for events lasting more than 1500 hours.

To assess the probability of lifetime to be  $\leq 1500$  hours we need to assess the probability of the union of events Sh and M, which is equal to the sum of their probabilities, as their intersection is equal to zero (disjoint events):

$$P(h \leq 1500) = P(\text{Sh} \cup \text{M}) = P(\text{Sh}) + P(\text{M}) = 0.225 + 0.4 = 0.625$$

Note that since events Sh, M and L are not only disjoint, but also exhaustive, they form a partition of sample space.

### CONCLUSION

In this first introductory article, we have briefly reviewed the current applications of Venn diagrams in

**Table 2. Lifetime of cells analysis example**

Таблица 2. Пример анализа продолжительности жизни клеток

Lifetime, h	Number of cells	Proportion
<1000	45	0.225
1000–1500	80	0.400
>1500	75	0.375

clinical research and shown that even such a classical way of visualization is still under development. We have then given several tutorials discussing the basic concepts of probability as well as its relation to the set theory and the application of the Venn diagrams for graphical

representation of these concepts. We understand that this material may give the impression of being purely abstract, however we stress that it is crucial to reach a good understanding of the principles of statistical analysis since Venn diagrams are often used in research papers.

### 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.

### 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 осуществлял руководство ее написанием. Все авторы участвовали в обсуждении и редактировании работы. Все авторы утвердили окончательную версию публикации.

### REFERENCES / ЛИТЕРАТУРА

- 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-fcaad2679185> 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.

### INFORMATION ABOUT THE AUTHORS / ИНФОРМАЦИЯ ОБ АВТОРАХ

**Nikolay M. Bulanov**<sup>✉</sup>, 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>

**Буланов Николай Михайлович**<sup>✉</sup>, 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>

**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>

**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>

✉ Corresponding author / Автор, ответственный за переписку