

МИНИСТЕРСТВО ЗДРАВООХРАНЕНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ
«ГОМЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ МЕДИЦИНСКИЙ УНИВЕРСИТЕТ»

Кафедра общественного здоровья и здравоохранения
с курсом ФПК и П

Н. В. ГАПАНОВИЧ-КАЙДАЛОВ, Т. М. ШАРШАКОВА

МЕДИЦИНСКАЯ СТАТИСТИКА **И ОСНОВЫ ДОКАЗАТЕЛЬНОЙ МЕДИЦИНЫ**

Учебно-методическое пособие
для студентов 4–6 курсов всех факультетов и магистрантов
учреждений высшего медицинского образования

Гомель
ГоГМУ
2020

УДК 61:311+614.1(072)

ББК 51:60.6я73

Г 19

Рецензент:

доктор медицинских наук, доцент,
заведующий кафедрой общественного здоровья и здравоохранения
Гродненского государственного медицинского университета

М. Ю. Сурмач;

доктор медицинских наук, доцент,
заведующий кафедрой общественного здоровья и здравоохранения
с курсом ФПК и П

Витебского государственного ордена Дружбы народов
медицинского университета

В. С. Глушанко

Гапанович-Кайдалов, Н. В.

Г 19 Медицинская статистика и основы доказательной медицины:
учеб.-метод. пособие для студентов 4–6 курсов всех факультетов и
магистрантов учреждений высшего медицинского образования /
Н. В. Гапанович-Кайдалов, Т. М. Шаршакова. — Гомель: ГомГМУ,
2020. — 44 с.

ISBN 978-985-588-192-7

Учебно-методическое пособие составлено для проведения практических за-
нятий по общественному здоровью и здравоохранению.

Предназначено для студентов 4–6 курсов всех факультетов и магистрантов
учреждений высшего медицинского образования.

Утверждено и рекомендовано к изданию научно-методическим советом
учреждения образования «Гомельский государственный медицинский университет»
20 февраля 2020 г., протокол № 3.

УДК 61:311+614.1(072)

ББК 51:60.6я73

ISBN 978-985-588-192-7

© Учреждение образования
«Гомельский государственный
медицинский университет», 2020

ОГЛАВЛЕНИЕ

Введение.....	4
1. Учебный материал	5
1.1. Основные понятия медицинской статистики и доказательной медицины	5
1.2. Методы медицинской статистики и доказательной медицины	16
1.3. Организация и проведение статистического исследования	25
2. Задания для самостоятельной работы студентов	29
3. Вопросы для самоконтроля.....	37
4. Тестовый контроль	38
Литература	41

ВВЕДЕНИЕ

Тема изучается в течение 6 часов и состоит из 3 частей.

В первой части занятия проводится изучение и обсуждение основных вопросов темы. Во второй части студенты под руководством преподавателя выполняют индивидуальные задания и решают задачи по теме занятия с целью отработки практических умений и навыков при расчетах статистических величин. Третья часть занятия посвящена самостоятельной работе студентов, контролю полученных знаний и практических навыков, подведению итогов и консультации преподавателя по содержанию рекомендованной литературы.

Мотивационная характеристика темы: статистический учет в здравоохранении дает развернутую характеристику состояния и развития отрасли, явлений и процессов в обществе в целом. Знания и умения, полученные при изучении данной темы, будут необходимы студентам при изучении ряда тем курса общественного здоровья и здравоохранения, в будущей научной деятельности и практической работе.

Цель занятия: углубить полученные теоретические знания о видах статистических величин, освоить методики расчета относительных и средних величин и навыки анализа статистического материала. Овладеть навыками расчета «риска» и оценки «шанса».

Отработать полученные умения и навыки при решении ситуационных задач. Выработать навыки использования статистических величин для оценки здоровья населения и деятельности системы здравоохранения.

Требования к исходному уровню знаний.

Студент должен знать:

- основы медицинской статистики;
- методы статистического анализа здоровья населения и деятельности организаций здравоохранения;
- виды статистических величин;
- методику расчета показателей (интенсивных, экстенсивных, соотношения, наглядности);
- средние величины и их применение в практической работе врача;
- определение терминов «риск», «шанс»;

Студент должен уметь:

- вычислять средние величины;
- рассчитывать относительные величины;
- вычислять среднеквадратичное отклонение;
- вычислять и оценивать критерии разнообразия признака;
- проводить оценку риска;
- вычислять и оценивать инцидентность;
- вычислять относительный и атрибутивный риск, добавочную долю популяционного риска.

1. УЧЕБНЫЙ МАТЕРИАЛ

1.1. Основные понятия медицинской статистики и доказательной медицины

Статистика — это вид практической деятельности, направленной на сбор, обработку, анализ и публикацию статистической информации, характеризующей количественные закономерности жизни общества (экономики, культуры, политики и др.); отрасль знаний (и соответствующие ей учебные дисциплины), в которой излагаются общие вопросы сбора, измерения и анализа массовых количественных данных.

Одной из отраслей статистики является **медицинская статистика**, которая изучает количественную сторону массовых явлений и процессов в медицине.

Статистика здоровья изучает здоровье общества в целом и его отдельных групп, устанавливает зависимость здоровья от различных факторов социальной среды.

Статистика здравоохранения анализирует данные о медицинских учреждениях, их деятельности, оценивает эффективность различных организационных мероприятий по профилактике и лечению болезней.

Доказательная медицина — это подход к медицинской практике, при котором решения о применении профилактических, диагностических и лечебных мероприятий принимаются исходя из имеющихся доказательств их эффективности и безопасности, а такие доказательства подвергаются поиску, сравнению, обобщению и широкому распространению для использования в интересах больных.

Критерии достоверности медицинской информации:

Высокая достоверность — информация основана на результатах нескольких независимых клинических испытаний с совпадением результатов, обобщенных в систематических обзорах.

Умеренная достоверность — информация основана на результатах по меньшей мере нескольких независимых, близких по целям клинических испытаний.

Ограниченная достоверность — информация основана на результатах одного клинического испытания.

Низкая достоверность — строгие научные доказательства отсутствуют (клинические испытания не проводились) — некое утверждение, основанное на мнении экспертов.

Признак — описание фактов, позволяющих согласно существующей теории сделать вывод о наличии интересующего явления.

Переменная — это количественно измеряемое свойство или признак, принимающий различные значения. Значения переменных могут изменяться либо непрерывно, либо дискретно.

Качественные данные выражают наличие или отсутствие у испытуемых того или иного признака, качества (боится ли ребёнок темноты, имеет ли испытуемый семью или высшее образование и т. п.).

Количественные данные — это наборы чисел, которые представляют собой некоторые величины, баллы, значения (например, балл по тесту, время реакции и т. п.).

Генеральная совокупность — это вся мыслимая совокупность измерений (например, результаты по тесту всех испытуемых, для которых этот тест предназначен).

Выборка — это часть генеральной совокупности ограниченного объёма, результаты измерения признака для некоторой группы объектов (испытуемых).

Репрезентативность (представительность) выборки — это её способность представлять генеральную совокупность (все категории объектов или испытуемых должны быть представлены в выборке).

Качественная однородность выборки: результаты обследования выборки испытуемых должны представлять собой некие индивидуальные величины одного и того же качества или свойства.

Однородность выборки (количественная однородность) — это различия между значениями признака. Чем больше диапазон (амплитуда) изменения значений признака, тем большей изменчивостью он обладает. Чем ближе друг к другу значения признака, тем меньше его изменчивость (вариативность).

Измерение — это приписывание числовых форм объектам или событиям в соответствии с определенными правилами.

Достоверность отдельных измерений. Чтобы обеспечить достоверность отдельных измерений некоторого показателя, используемый метод его измерения должен отражать все важные аспекты изучаемого явления и отображать меру согласованности результатов соответствующего теста с другими оценками того же явления, а также степень предсказуемости данных, полученные в ходе измерений.

Воспроизводимость результатов измерений определяется вероятностью получения одного и того же результата измерения показателя при повторных измерениях, сделанных разными людьми, на разных приборах, в разное время и в разных местах. В частности, воспроизводимость биохимических показателей можно оценить путем повторных измерений, выполненных на одном и том же образце, но разными специалистами или на разных приборах.

Точность измерений — это способность метода реагировать на изменения измеряемого показателя. Менее «точные» (т. е. более «грубые») методы, шкалы измерения которых имеют всего 3–5 различных числовых оценок, зачастую не позволяют выявить небольшие изменения тяжести со-

стояния больного, даже если сами больные уверенно констатируют наличие улучшения или ухудшения. Например, в процессе лечения больного ишемической болезнью сердца (ИБС) его состояние улучшается (уменьшается частота приступов стенокардии напряжения, увеличивается толерантность к физической нагрузке), однако у него все еще сохраняется тот же функциональный класс ИБС. Аналогично ведут себя такие показатели, как степень артериальной гипертензии или класс тяжести сердечной недостаточности: их шкалы измерения таковы, что соответствующие оценки запаздывают при выявлении динамики состояния больного на фоне лечения.

Интерпретируемость результатов зачастую может быть улучшена, если перейти от более детальных шкал оценки показателя к более грубым. Примеры: количественные значения уровней тревожности и депрессивности по шкалам опросника HADS не всегда понятны врачам и пациентам, что вызывает необходимость «огрублять» эти шкалы для получения оценки показателей как клинически значимых, субклинических или нормальных. Именно для улучшения интерпретируемости результатов измерений (и не только при применении опросников, но и при физикальном обследовании или выполнении лабораторных тестов) предлагаются границы нормы изучаемых показателей и т. п.

Шкала измерения — это форма фиксации совокупности признаков изучаемого объекта с упорядочиванием их в определённую числовую систему; метрические системы, моделирующие исследуемый феномен, объекты или события путём замены прямых обозначений изучаемых объектов числовыми в соответствии с определёнными правилами.

Дихотомическая шкала — номинативная шкала, состоящая из двух ячеек, например: «имеет братьев и/или сестёр — единственный ребёнок в семье»; «иностранец — соотечественник»; «имеет высшее образование — не имеет высшего образования» и т. п.

Графическая шкала — это форма фиксации данных при помощи наглядного отображения развития признака в виде непрерывной линии или определённой фигуры.

Номинативная шкала — это шкала, классифицирующая по названию: *poter* (лат.) — имя, название. Название же не измеряется количественно, оно лишь позволяет отличить один объект от другого или одного субъекта от другого. Номинативная шкала — это способ классификации объектов или субъектов, распределения их по ячейкам классификации.

Интервальная шкала — это шкала, классифицирующая по принципу «больше на определенное количество единиц — меньше на определенное количество единиц». Каждое из возможных значений признака отстоит от другого на равном расстоянии.

Шкала равных отношений — это шкала, классифицирующая объекты или субъектов пропорционально степени выраженности измеряемого

свойства. В шкалах отношений классы обозначаются числами, которые пропорциональны друг другу: 2 так относится к 4, как 4 к 8. Это предполагает наличие абсолютной нулевой точки отсчета.

Порядковая шкала — это шкала, классифицирующая по принципу «больше — меньше». Если в шкале наименований было безразлично, в каком порядке мы расположим классификационные ячейки, то в порядковой шкале они образуют последовательность от ячейки «самое малое значение» к ячейке «самое большое значение» (или наоборот).

Первичная обработка данных — упорядочивание, ранжирование, группировка, представление в табличном виде; вычисление мер центральной тенденции (моды, медианы, среднего значения), мер изменчивости (размаха, дисперсии, стандартного отклонения); наглядное представление данных в виде графиков и диаграмм.

Вторичная статистическая обработка — выдвижение статистических гипотез, подготовка данных для применения статистических методов (проверка нормальности распределения, создание и уравнивание дисперсионных комплексов, проверка достоверности экстремальных значений и др.), проверка гипотез с помощью выбранных статистических критериев, формулирование выводов, имеющих определенную доверительную вероятность.

Статистические величины (показатели) подразделяются на абсолютные, относительные и средние (рисунок 1).

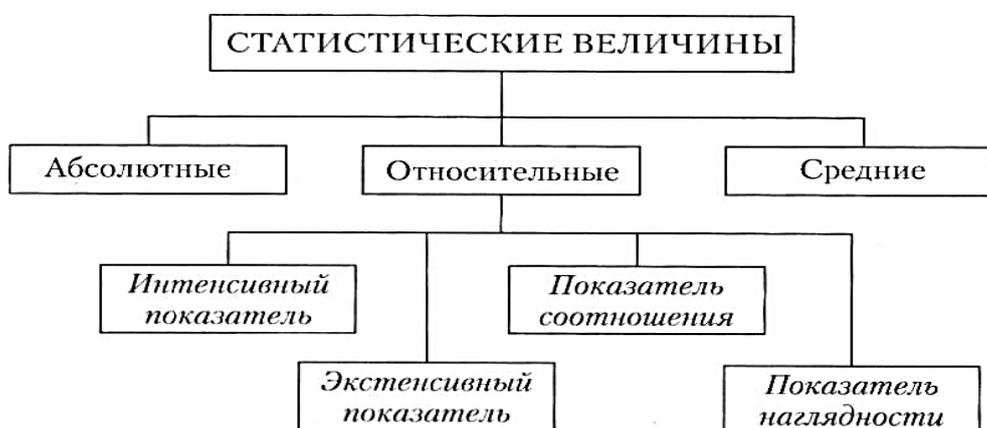


Рисунок 1 — Основные виды статистических величин

Абсолютные статистические показатели характеризуют абсолютные размеры изучаемых статистикой процессов и явлений: их массу, площадь, объем, протяженность; отражают их временные характеристики, а также могут представлять объем совокупности, т. е. число составляющих ее единиц.

Относительные статистические показатели (относительные статистические величины) представляют собой частное от деления двух

статистических величин и характеризуют количественное соотношение между ними.

Интенсивные показатели — показатели частоты, интенсивности, распространенности явления в среде, продуцирующей данное явление.

$$\text{Интенсивный показатель} = \frac{\text{Абсолютный размер явления}}{\text{Абсолютный размер среды, продуцирующий данное явление}} \times \text{основание.}$$

Величина основания выбирается в соответствии с величиной показателя — на 100, 1000, 10000, 100000, в зависимости от этого показатель выражается в *процентах, промилле, продецимилле, просантимилле*.

Интенсивные показатели могут быть: *общими* — характеризуют явление в целом (общие показатели рождаемости, смертности, заболеваемости, вычисленные ко всему населению административной территории); *специальными (погрупповыми)* — применяются для характеристики частоты явления в различных группах (заболеваемость по полу, возрасту, смертность среди детей в возрасте до 1 года, летальность по отдельным нозологическим формам и т. д.).

Интенсивные показатели применяются в медицине: для определения уровня, частоты, распространенности явления, для сравнения частоты явления в двух различных совокупностях, для изучения изменений частоты явления в динамике.

Экстенсивные показатели — показатели удельного веса, структуры, характеризуют распределение явления на составные части, его внутреннюю структуру. Вычисляются экстенсивные показатели отношением части явления к целому и выражаются в процентах или долях единицы.

$$\text{Экстенсивный показатель} = \frac{\text{Абсолютный размер части явления}}{\text{Абсолютный размер явления в целом}} \times 100.$$

Экстенсивные показатели используются для определения структуры явления и сравнительной оценки соотношения составляющих его частей. Экстенсивные показатели всегда взаимосвязаны между собой, т. к. их сумма всегда равна 100 процентам: так, при изучении структуры заболеваемости удельный вес отдельного заболевания может возрасти: при истинном росте числа заболеваний; при одном и том же его уровне, если число других заболеваний снизилось; при снижении числа данного заболевания. Если уменьшение числа других заболеваний происходит более быстрыми темпами.

При анализе экстенсивный показатель следует применять с осторожностью и помнить, что им пользуются только для характеристики состава (структуры) явления в данный момент времени и в данном месте.

Примеры использования в работе врача: лейкоцитарная формула; структура населения по полу, возрасту, социальному положению; структура заболеваний по нозологии; структура причин смерти.

Показатели соотношения — представляют собой соотношение двух самостоятельных, независимых друг от друга, качественно разнородных величин, сопоставляемых только логически.

Показатель соотношения = $\frac{\text{Абсолютный размер явления}}{\text{Абсолютный размер среды, непродуцирующей данное явление}} \times \text{основание}$.

Примеры использования в работе врача: показатели обеспеченности населения врачами, больничными койками; показатели, отражающие число лабораторных исследований на 1 врача и т. д.

Показатели наглядности — применяются с целью более наглядного и доступного сравнения статистических величин. Показатели наглядности представляют удобный способ преобразования абсолютных, относительных или средних величин в легкую для сравнения форму. При вычислении этих показателей одна из сравниваемых величин приравнивается к 100 (или 1), а остальные величины пересчитываются соответственно этому числу.

Показатель наглядности = $\frac{\text{Явление}}{\text{Такое же явление из ряда сравниваемых, принятое за 100}} \times 100$.

Показатели наглядности указывают, на сколько процентов или во сколько раз произошло увеличение или уменьшение сравниваемых величин. Показатели наглядности используются чаще всего для сравнения данных в динамике, чтобы представить закономерности изучаемого явления в более наглядной форме.

Основными формами наглядного представления данных исследования являются: таблицы, диаграммы, графики.

Таблицы представляют собой упорядоченные по горизонтали и по вертикали наборы количественных и качественных данных, заключённые в рамки или без них.

График — это некоторая линия на плоскости (или в пространстве), отображающая зависимость между двумя (тремя) переменными. По оси абсцисс откладывают значения независимой переменной, а по оси ординат — зависимой переменной.

Диаграммы — это способ графического изображения величин при помощи фигур (столбцов, секторов и т. п.), площади которых пропорциональны этим величинам; используются, главным образом, для изображения соотношений между величинами.

Круговая (секторная) диаграмма — наглядное изображение данных, выраженных в процентах, в виде круговых секторов, имеющих соответствующие площади.

Вероятность — степень возможности проявления какого-либо определённого события в тех или иных условиях.

Кумулятивная вероятность выживания в момент времени t — это то же самое, что доля выживших пациентов к этому моменту времени.

Шанс — это отношение вероятности того, что события произойдет к вероятности того, что событие не произойдет:

1) если шанс = 1, то вероятность наступления события равна вероятности того, что событие не произойдет;

2) если шанс >1, то вероятность наступления события больше вероятности того, что событие не произойдет;

3) если шанс <1, то вероятность наступления события меньше вероятности того, что событие не произойдет.

Отношение шансов — это отношение шансов для первой группы объектов к отношению шансов для второй группы объектов:

1) если отношение шансов = 1, то шанс для первой группы равен шансу для второй группы;

2) если отношение шансов >1, то шанс для первой группы больше шанса для второй группы;

3) если отношение шансов <1, то шанс для первой группы меньше шанса для второй группы.

Средние величины — это величины, вокруг которых группируются остальные данные; обобщают всю выборку, позволяют судить обо всей выборке, дают возможность сравнивать разные выборки, разные серии между собой виды: выборочное среднее, медиана, мода.

Выборочное среднее — это результат деления суммы всех значений (X) на их количество (N):

$$\bar{x} = \frac{\sum x}{n} .$$

Медиана — это значение изучаемого признака, которое делит выборку, упорядоченную по величине данного признака, пополам.

Мода — это значение, наиболее часто встречающееся в выборке, т. е. значение с наибольшей частотой. Если все значения в группе встречаются одинаково часто, то считается, что моды нет.

Меры изменчивости — это статистические показатели, характеризующие различия между отдельными значениями выборки: размах, дисперсия, стандартное отклонение.

Размах — это разность между наибольшим и наименьшим из отмеченных или возможных значений переменной.

Дисперсия характеризует отклонения от средней величины в данной выборке:

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n} .$$

Стандартное отклонение:

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

Коэффициент вариации — это процентное отношение среднеквадратического отклонения к среднеарифметической величине, относительная мера колеблемости вариационного ряда. При $C_v < 10\%$ разнообразие ряда считается слабым, при C_v от 10 до 20 % — средним, а при $C_v > 20\%$ — сильным:

$$C_v = \frac{\sigma}{\bar{x}} \cdot 100\%$$

Распределением признака называется закономерность встречаемости разных его значений.

Параметры распределения — это его числовые характеристики, указывающие, где «в среднем» располагаются значения признака, насколько эти значения изменчивы и наблюдается ли преимущественное появление определенных значений признака.

Ряд распределения — разновидность структурной группировки по одному признаку, отображенная в групповой таблице с графами, в которых содержатся варианты, абсолютные и/или относительные частоты признака (таблица 1).

Таблица 1 — Ряд распределения случайной величины

Значение	f_a (абсолютная частота)	f_o (относительная частота)
1	3	0,05
2	12	0,20
3	21	0,35
4	15	0,25
5	9	0,15
Σ (сумма)	60	1

Абсолютная (f_a) и относительная (f_o) частоты связаны соотношением:

$$f_o = \frac{f_a}{n},$$

где f_a — абсолютная частота, f_o — относительная частота, n — объём выборки.

Сумма абсолютных частот равна объёму выборки ($\Sigma f_a = n$).

Сумма относительных частот равна единице ($\Sigma f_o = 1$).

Интервальный вариационный ряд — упорядоченная совокупность интервалов варьирования значений признака с соответствующими частотами попаданий в каждый из них значений величины (таблица 2).

Таблица 2 — Интервальное распределение случайной величины

Интервал	f_a	f_o
32–35	1	0,025
36–39	3	0,075
40–43	4	0,1
44–47	5	0,125
48–51	9	0,225
52–55	7	0,175
56–59	6	0,15
60–63	3	0,075
64–67	2	0,05
Σ	40	1

Динамический ряд — ряд однородных сопоставимых величин, показывающих изменение изучаемого явления во времени (отображение развития явления во времени). Числа, составляющие динамический ряд, принято называть уровнями ряда. *Уровни ряда могут быть представлены абсолютными числами, относительными и средними величинами.*

Для более глубокого понимания сути развития общественных явлений исчисляют такие показатели динамического ряда, как абсолютный прирост, темп прироста, темп роста, абсолютное значение 1 % прироста.

Абсолютным приростом называют разницу между каждым последующим уровнем и уровнем предыдущим. Абсолютный прирост может быть положительным и отрицательным.

Темпом роста называется отношение каждого последующего уровня к предыдущему, выраженному в процентах.

Темпом прироста называется отношение абсолютного прироста к предыдущему уровню, принятому за 100 %.

Так как каждому относительному показателю соответствуют определенные абсолютные величины, то при изучении темпов прироста нужно обязательно учитывать, какая абсолютная величина соответствует каждому проценту прироста, каково его содержание. Для этого исчисляется такой показатель, как *абсолютное значение одного процента прироста*. Он определяется как частное от деления абсолютного прироста за определенный период на темп прироста в процентах за этот же период.

Для иллюстрации расчетов рассмотренных статистических показателей приведем ряд динамики.

Гистограмма — это графическое представление частотного распределения, столбчатая диаграмма, состоящая из вертикальных прямоугольников, основания которых расположены на одной прямой (рисунок 2). По оси

абсцисс откладывают значения наблюдаемой величины, а по оси ординат — абсолютные или относительные частоты встречаемости отдельных значений (или количество значений, попадающих в определённые интервалы).

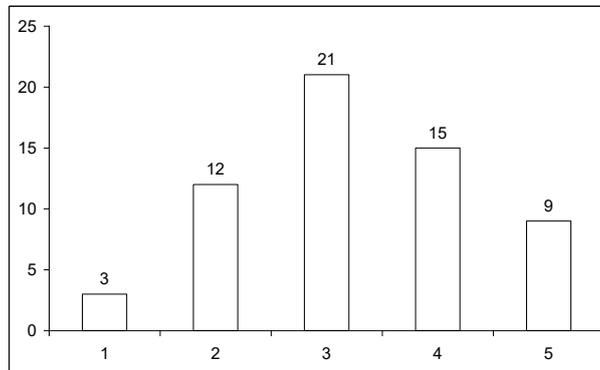


Рисунок 2 — Гистограмма на основе частотного распределения случайной величины

Полигон частот — это графическое представление частотного распределения в результате соединения на графике точек отрезками прямой (рисунок 3). По своей форме этот график представляет многостороннюю фигуру (полигон).



Рисунок 3 — Полигон частот распределения случайной величины

Нормальное распределение имеет наиболее широкое применение в математической статистике (рисунок 4). К нормальному распределению близки распределения ошибок измерений, отклонения от нормы (среднего) значения уровня интеллекта и т. д.

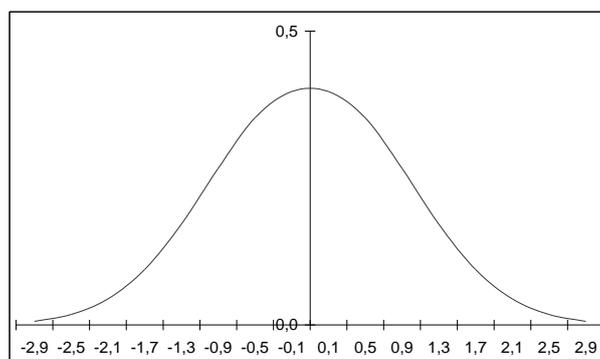


Рисунок 4 — Кривая нормального распределения

Относительный риск рассчитывается как отношение риска развития заболевания в «экспонированной» группе (F^+) к риску развития заболевания в «неэкспонированной» группе (F^-):

$$RR = \frac{\text{частота случаев в группе } F^+}{\text{частота случаев в группе } F^-} .$$

Относительный риск демонстрирует, во сколько раз действие фактора риска увеличивает вероятность развития заболевания. Значение отношения рисков, равное 1, говорит о том, что фактор риска не оказывает влияние на развитие заболевания. Если относительный риск больше 1, то можно сделать вывод о том, что действие изучаемого фактора увеличивает риск развития заболевания, и чем больше значение относительного риска, тем выше вероятность его развития. Если относительный риск меньше 1, то фактор, наоборот, снижает вероятность развития заболевания.

Атрибутивный риск рассчитывается как разница между вероятностью развития заболевания в «экспонированной» (F^+) и «неэкспонированной» (F^-) группе:

$$AR = \text{частота случаев в группе } F^+ - \text{частота случаев в группе } F^- .$$

На практике данный показатель демонстрирует, какую заболеваемость создает действие данного фактора риска.

Доля атрибутивного риска представляет собой атрибутивный риск в процентном выражении.

Для использования данных показателей в отношении всей популяции можно рассчитать: **добавочный популяционный риск**, рассчитывается как произведение атрибутивного риска на распространенность фактора риска в популяции (позволяет оценить заболеваемость в популяции, связанную с распространением фактора риска); **добавочную долю популяционного риска (этиологическую долю)**, рассчитывается как отношение добавочного популяционного риска к общей заболеваемости в популяции.

Компьютерная обработка статистических данных производится с помощью специальных пакетов прикладных программ, обеспечивающих быстроту статистических расчетов, высокую надежность и достоверность результатов, возможность представления данных в аналитической, графической или табличной формах.

Примерами таких программ являются:

1. *Microsoft Excel* подходит для накопления данных, промежуточного преобразования, предварительных статистических прикидок, для построения некоторых видов диаграмм.

2. *SPSS (Statistical Package for Social Science)* позволяет вводить и редактировать данные, представлять их в виде наглядных отчетов, содержит

в себе более 100 процедур статистического анализа, а также систему помощи и подсказок.

3. *STATISTICA* включает большое количество методов статистического анализа, объединенных следующими специализированными статистическими модулями:

- Основные статистики и таблицы.
- Непараметрическая статистика.
- Дисперсионный анализ.
- Множественная регрессия.
- Нелинейное оценивание.
- Анализ временных рядов и прогнозирование и др.

1.2. Методы медицинской статистики и доказательной медицины

В медицинских исследованиях наиболее часто используют две группы методов:

1. Критерии, позволяющие выявить различия между исследуемыми препаратами и изучаемыми группами больных по каким-либо параметрам.

2. Методы, устанавливающие наличие и степень взаимосвязи между двумя или несколькими признаками.

Реже используется третья группа, к которой относятся более сложные способы анализа данных — многомерные методы исследования (факторный анализ, кластерный анализ и т. д.).

В первую очередь выбор критериев определяется тем, какие параметры предполагается анализировать:

— ***количественные*** (например, уровень артериального давления и т. д.), их можно упорядочить и производить над ними арифметические действия;

— ***порядковые*** (например, степень тяжести заболевания — легкая, средняя, тяжелая) их можно упорядочить, но арифметические действия над ними производить нельзя;

— ***качественные*** (например, окраска бактерий по Граму — красная или синяя), их нельзя упорядочить и над ними нельзя производить арифметические действия, их можно описать, подсчитав число наблюдений, имеющих одно и то же значение, а затем оценить (сравнить) группы по количеству наблюдений, имеющих различные значения.

Методические проблемы определения границ нормы. Следует подчеркнуть: определение того, что можно считать нормой для некоторого показателя в конкретном клиническом или эпидемиологическом исследовании, не свободно от серьезных методических проблем.

Иногда можно встретить определения, непосредственно основанные на результатах применения статистических методик к выборочным рас-

пределениям показателя. Например, в качестве границы нормы используется пороговая точка для верхней квинтили (т. е. 80 %-ная отрезная точка) или верхней квартили распределения (т. е. 75 %-ная отрезная точка). Однако, если бы все значения показателей, находящиеся за некоторым статистическим пределом (скажем, за 80 %-ной отрезной точкой), рассматривались в качестве патологических, то и распространенность всех соответствующих патологий оказалась бы одинаковой и составляла точно 20 %.

Кроме того, известно, что риск наличия заболевания коррелирует с показателями многих клинико-лабораторных тестов по всему диапазону их возможных значений (т. е. при разных пороговых значениях «нормы»). Например, в пределах нормальных значений уровня общего холестерина риск заболевания ИБС различается в несколько раз, если сравнивать «высокие нормальные» и «низкие нормальные» значения этого показателя. Аналогичная ситуация наблюдается и в случае других факторов риска сердечно-сосудистых заболеваний (ССЗ) и смерти от ССЗ: например, для уровня систолического и диастолического артериального давления и т. п.

Таким образом, фактически не существует однозначного соответствия между степенью статистической «необычности» измеренного значения показателя и наличием соответствующего заболевания. Более того, некоторые крайние, явно необычные значения показателя могут — в случае конкретного больного — оказаться для него предпочтительнее «обычных» значений этого показателя.

Естественно, следует обращать особое внимание на то, являются ли подобные необычные результаты измерения воспроизводимыми у данного больного.

Предупреждение систематических и случайных ошибок измерений. В статистическом смысле, систематическая ошибка измерения показателя (иными словами — смещение) — это некоторое однонаправленное отклонение результатов измерения показателя, т. е. его систематическое завышение или занижение. При этом величина систематической ошибки не зависит от объема выборки.

Как правило, систематическая ошибка может быть оценена и даже уменьшена еще на этапе планирования исследования (с помощью стандартизации измерений, калибровки приборов и т. п.) и учтена впоследствии — при анализе результатов.

Кроме того, основные источники систематических ошибок в клинических исследованиях хорошо известны. Прежде всего, это погрешности выборки, когда организатор исследования был необъективен при распределении обследуемых по группам сравнения, а также ошибки измерения, когда в одной из групп пациенты лучше определяют последствия лечения, чем в другой.

Здесь играют свою роль и ошибки памяти, из-за которых испытуемые в одной группе точнее помнят отдельные события, чем в другой, а также

эффект плацебо, т. е. положительный эффект приема «пустышки» у тех испытуемых, которые думают, что они принимают активный препарат.

Метод когортных исследований. Вместо измерения вклада в уже существующее заболевание (или его отсутствие), как это делается в одномоментном исследовании, когортные исследования изучают вклад факторов в развитие или прогрессирование заболевания. Когорта — римский термин для группы солдат, которые маршировали вместе. В клиническом исследовании когорта — это группа обследуемых, прослеженных в течение времени. В когортном исследовании исследователь выбирает или формирует выборку пациентов (обследуемых), а также измеряет показатели (переменные, факторы) у каждого обследуемого, например, физическая активность, которые могут повлиять на исходы.

Когда когорта формируется для изучения, в основном, одного заболевания (или фактора риска), исследователи часто исключают людей, которые уже имеют событие (и это называется *inception cohort*). При исключении лиц, о которых известно, что они имеют уже интересующий исход (в данном случае, заболевание или фактор риска), исследователь уверен, что переменные, которые измеряются в начале исследования, не влияют на исходы. Однако некоторые состояния могут присутствовать и продуцировать симптомы до постановки диагноза. Потенциальные проблемы, связанные с тем, что некий фактор может влиять на возникновение события, например, питание, могут быть минимизированы двумя путями. Во-первых при скрининге должны быть использованы чувствительные тесты и исключены потенциальные обследуемые с субклиническими формами интересующей болезни. Во-вторых — исследователь может увеличить временные рамки, спрашивая на скрининге о прошлых привычках питания или увеличивая длительность наблюдения так, чтобы период от измерения фактора риска до возникновения события был дольше, чем преклиническая фаза болезни.

Метод выявления взаимосвязи (корреляции). Известны два вида связи между явлениями (признаками): функциональная и корреляционная.

Функциональные связи — это связи, при которых определенному значению одной переменной (признака-фактора) соответствует однозначно определяемое значение другой переменной (признака-результата). Функциональная зависимость предполагает, что уровень признака-результата полностью определяется величиной признака-фактора (факторов). Если между признаками установлена функциональная зависимость, то она справедлива для каждой единицы статистической совокупности.

Статистические связи характеризуются тем, что определенному значению одной переменной (признака-фактора) соответствует распределение (множество) значений другой переменной (признака-результата). Частным случаем статистической связи является корреляционная зависимость, при которой одному значению признака-фактора соответствует множество

значений признака-результата, причём закономерным образом изменяется средняя величина результативного признака. Таким образом, при изучении статистической (корреляционной) зависимости выявляется только тенденция изменения признака-результата под влиянием изменения признака-фактора (факторов), статистические связи могут не проявляться на уровне отдельно взятой единицы изучаемой совокупности. Социально-экономическим явлениям присущи, в основном, статистические связи, поскольку их развитие — это результат одновременного воздействия большого числа причин.

Корреляционная связь необходима, например, при оценке взаимосвязей между стажем работы и уровнем заболеваемости работающих; между разными уровнями физических факторов окружающей среды и состоянием здоровья; между различными уровнями интенсивности нагрузки и частотой (уровнем) физиологических реакций организма; между сроками госпитализации и частотой осложнений.

Степень тесноты связи (линейная форма) можно измерить с помощью линейного коэффициента корреляции:

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{(n \sum x^2 - (\sum x)^2)(n \sum y^2 - (\sum y)^2)}}$$

где n — количество пар значений двух признаков X и Y .

Степень тесноты связи можно также измерить с помощью ранговой корреляции Спирмена:

$$R_S = 1 - \frac{6 \sum (R_x - R_y)^2}{n^3 - n}$$

где n — количество пар значений двух признаков X и Y ; R_x и R_y — ранги соответствующих значений двух признаков.

Таблица 3 — Оценка степени тесноты и характера связи

R	Интерпретация
около 0	отсутствие связи
до 0,29	слабая
0,3–0,69	средняя
0,7–0,99	сильная
1,0	полная
+	прямая
–	обратная

Регрессионный анализ

При наличии корреляционной связи между факторными и результативными признаками врачам нередко приходится устанавливать, на какую величину может измениться значение одного признака при изменении другого на общепринятую или установленную самим исследователем единицу измерения.

Общий вид корреляционного уравнения прямой линии регрессии, т. е. уравнения прямолинейной корреляционной связи, выглядит следующим образом:

$$y = a + b \times x,$$

где x и y — индивидуальные значения соответственно факториального и результативного признаков; a и b — параметры уравнения прямолинейной и корреляционной связи.

$$b = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2};$$

$$a = \bar{y} - b \cdot \bar{x}.$$

1.3. Организация и проведение статистического исследования

Статистическое исследование позволяет получить представление о том или ином явлении, изучить его размер, уровень, выявить закономерности. Предметом исследования могут быть здоровье населения, организация медицинской помощи, факторы внешней среды, влияющие на здоровье и т. д.

1. Изучение тенденций состояния здоровья населения проводится на генеральных совокупностях или достаточно больших по численности выборочных совокупностях, позволяющих получить интенсивные показатели и обоснованно перенести полученные данные на всю генеральную совокупность.

2. Проведение строго спланированных исследований по изучению отдельных факторов без выявления интенсивности явления в среде — проводится, как правило, на небольших по численности совокупностях с целью выявления новых факторов, изучения неизвестных или малоизвестных причинно-следственных связей.

Этапы статистического исследования:

1 этап. Составление плана и программы исследования — является подготовительным, на нем определяется цель и задачи исследования, составляется план и программа исследования, разрабатывается программа сводки статистического материала и решаются организационные вопросы.

А. Цель и задачи исследования должны быть четко сформулированы; цель определяет основное направление исследования и носит, как правило, не только теоретический, но и практический характер, она формулируется ясно, четко, недвусмысленно; для раскрытия поставленной цели определяются задачи исследования.

Б. Необходимо изучить по данной теме литературу.

В. Необходимо разработать «Организационный план» — предусматривает определение: 1) места (административно-территориальных границ наблюдения); 2) времени (конкретных сроков осуществления наблюдения, проведения разработки и анализа материала); 3) субъекта исследования (организаторов, исполнителей, методического и организационного руководства, источников финансирования исследования).

Г. Разработка «Плана исследования» — включает определение:

- объекта исследования (статистической совокупности);
- объема исследования (сплошное, несплошное);
- видов (текущее, единовременное);
- способов сбора статистической информации.

Д. Необходимо составить «Программу исследования (наблюдения)» — включает:

- определение единицы наблюдения;
- перечень вопросов (учетных признаков), подлежащих регистрации в отношении каждой единицы наблюдения;
- разработку индивидуального учетного (регистрационного) бланка с перечнем вопросов и признаков, подлежащих учету;
- разработку макетов таблиц, в которые затем вносятся результаты исследования.

На каждую единицу наблюдения заполняется отдельный бланк, он содержит паспортную часть, четко сформулированные, поставленные в определенной последовательности вопросы программы и дату заполнения документа. В качестве учетных бланков могут быть использованы применяемые в практике лечебно-профилактических учреждений учетные медицинские формы.

Источниками получения информации могут служить другие медицинские документы (истории болезни, и индивидуальные карты амбулаторного больного, истории развития ребенка, истории родов), отчетные формы лечебно-профилактических учреждений и др.

Для обеспечения возможности статистической разработки данных из этих документов производят выкопировку сведений на специально разработанные учетные бланки, содержание которых определяется в каждом отдельном случае в соответствии с задачами исследования.

В настоящее время в связи с машинной обработкой результатов наблюдения с использованием ЭВМ вопросы программы могут быть формализованы, когда вопросы в учетном документе ставятся в виде альтерна-

тивы (да, нет), или предлагаются уже готовые ответы, из которых следует выбрать определенный ответ.

Е. Необходимо составить программу сводки полученных данных, которая включает установление принципов группировки, выделение группировочных признаков, определение комбинаций этих признаков, составление макетов статистических таблиц.

2 этап. Сбор материала (статистическое наблюдение) — заключается в регистрации отдельных случаев изучаемого явления и характеризующих их учетных признаков в регистрационные бланки. Перед и в ходе выполнения этой работы проводится инструктаж (устный или письменный) исполнителей наблюдения, обеспечение их формами регистрации.

Статистическое наблюдение может быть:

А. По времени:

1. **Текущим** — явление изучается за какой-то отдельный период времени (неделю, квартал, год и т. д.) путем повседневной регистрации явления по мере возникновения каждого случая (учет числа родившихся, умерших, заболевших, выписанных из стационара). Так учитываются быстро меняющиеся явления.

2. **Единовременным** — статистические данные собираются на определенный (критический) момент времени (перепись населения, изучение физического развития детей, профилактические осмотры населения). Единовременная регистрация отражает состояние явления на момент изучения, используется для изучения медленно меняющихся явлений.

Выбор вида наблюдения по времени определяется целью и задачами исследования (характеристику госпитализированных больных можно получить в результате текущей регистрации выбывших из стационара — текущее наблюдение или путем однодневной переписи больных, находящихся в стационаре — единовременное наблюдение).

Б. В зависимости от полноты охвата изучаемого явления:

1. **Сплошное** — изучаются все входящие в состав совокупности единицы наблюдения, т. е. генеральная совокупность. Проводят с целью установления абсолютных размеров явления (общей численности населения, общего количества родившихся или умерших). Применяется и в тех случаях, когда сведения необходимы для оперативной работы (учет инфекционной заболеваемости, нагрузка врачей и др.).

2. **Несплошное** — изучается лишь часть генеральной совокупности, делится на несколько видов:

- **Монографический метод** — дает детальное описание отдельных характерных в каком-либо отношении единиц совокупности и глубокое, всестороннее описание объектов.

- **Метод основного массива** — предполагает изучение тех объектов, в которых сосредоточено значительное большинство единиц наблюдения.

Недостатком этого метода является то, что остается неохваченной исследованием часть совокупности, хотя и небольшая по размерам, но которая может значительно отличаться от основного массива.

• **Анкетный метод** — это сбор статистических данных с помощью специально разработанных анкет, адресованных определенному кругу лиц. Это исследование основано на принципе добровольности, поэтому возврат анкет зачастую бывает неполным. Нередко ответы на поставленные вопросы носят отпечаток субъективности и случайности. Этот метод применяется для получения приблизительной характеристики изучаемого явления.

• **Выборочный метод** — самый распространенный метод, сводится к исследованию некоторой специально отобранной части единиц наблюдения для характеристики всей генеральной совокупности. Преимуществом этого метода является получение результатов высокой степени надежности, а также значительно более низкая стоимость. В исследовании занято меньшее число исполнителей, кроме того он требует меньших затрат времени. В медицинской статистике роль и место выборочного метода особенно велики, поскольку медицинские работники имеют дело обычно только с частью изучаемого явления (изучают группу больных с тем или иным заболеванием, анализируют работу отдельных подразделений).

В. По способу получения сведений в ходе проведения и характеру его осуществления:

1. **Непосредственное наблюдение** (клинический осмотр больных, проведение лабораторных, инструментальных исследований, антропометрические измерения и т. п.).

2. **Социологические методы:** метод интервью (очный опрос), анкетирование (заочный опрос — анонимный или неанонимный) и др.

3. **Документальное исследование** (выкопировка сведений из учетно-отчетных медицинских документов, сведения официальной статистики учреждений и организаций).

3 этап. Разработка материала, статистическая группировка и сводка — начинается с проверки и уточнения числа наблюдений, полноты и правильности полученных сведений, выявлении и устранении ошибок, дубликатов записей и т. д.

Для правильной разработки материала применяется шифровка первичных учетных документов, т. е. обозначение каждого признака и его группы знаком — буквенным или цифровым. Шифровка — это технический прием, облегчающий и ускоряющий разработку материала, повышающий качество, точность разработки. Шифры — условные обозначения — вырабатываются произвольно. При шифровке диагнозов рекомендуется пользоваться международной номенклатурой и классификацией болезней; при шифровке профессий — словарем профессий.

Преимуществом шифровки является то, что при необходимости после окончания основной разработки можно вернуться к материалу для разра-

ботки с целью выяснения новых связей и зависимостей. Зашифрованный учетный материал позволяет сделать это легче и быстрее, чем незашифрованный. После проверки проводится группировка признаков.

Группировка — расчленение совокупности изучаемых данных на однородные типичные группы по наиболее существенным признакам. Группировка может проводиться по качественным и количественным признакам. Выбор группировочного признака зависит от характера изучаемой совокупности и задач исследования.

А. Типологическая группировка производится по качественным (описательным, атрибутивным) признакам (пол, профессия, группы болезни).

Б. Вариационная группировка (по количественным признакам) проводится на основании числовых размеров признака (возраст, длительность заболевания, продолжительность лечения и т. д.). Количественная группировка требует решения вопроса о величине группировочного интервала: интервал может быть равным, а в ряде случаев — неравным, даже включать так называемые открытые группы (при группировке по возрасту могут быть определены открытые группы: до 1 года, 50 лет и старше).

При определении числа групп исходят из цели и задач исследования. Необходимо, чтобы группировки могли вскрыть закономерности изучаемого явления. Большое число групп может привести к чрезмерному дроблению материала, ненужной детализации. Малое число групп приводит к затушевыванию характерных черт.

Рекомендуется следующее число групп: до 40 наблюдений — 5–6 групп, 40–60 наблюдений — 6–8 групп, 60–100 наблюдений — 7–10 групп и т. д.

Закончив группировку материала, приступают к **сводке** — обобщению единичных случаев, полученных в результате статистического исследования, в определенные группы, их подсчет и внесение в макеты таблиц.

Сводку статистического материала проводят при помощи статистических таблиц. Таблица, не заполненная цифрами, называется **макетом**.

Статистические таблицы бывают перечневые, хронологические, территориальные.

Таблица имеет подлежащее и сказуемое. Статистическое подлежащее обычно размещается по горизонтальным строкам в левой части таблицы и отражает главный, основной признак. Статистическое сказуемое размещается слева направо по вертикальным графам и отражает дополнительные учетные признаки.

Статистические таблицы делятся на:

А. **Простые** — представлено числовое распределение материала по одному признаку, составных частей его. Простая таблица содержит обычно простой перечень или итог по всей совокупности изучаемого явления.

Б. **Групповые** — представлено сочетание двух признаков в связи друг с другом.

В. Комбинационные — дается распределение материала по трем и более взаимосвязанным признакам.

При составлении таблиц должны соблюдаться определенные требования:

— каждая таблица должна иметь заголовок, отражающий ее содержание;
— внутри таблицы все графы также должны иметь четкие краткие названия;

— при заполнении таблицы все клетки таблицы должны содержать соответствующие числовые данные. Оставшиеся незаполненными из-за отсутствия данной комбинации клетки таблицы прочеркиваются («—»), а при отсутствии сведений в клетке проставляется «н. с.» или «...»;

— после заполнения таблицы в нижней горизонтальном ряду и в последнем справа вертикальном столбце подводятся итоги вертикальных граф и горизонтальных строк;

— таблицы должны иметь единую последовательную нумерацию.

В исследованиях, имеющих небольшой объем наблюдений, сводка проводится вручную. Все учетные документы раскладываются на группы в соответствии с шифром признака. Далее проводится подсчет и запись данных в соответствующую клетку таблицы. В настоящее время в проведении сортировки и сводки материала широко используются ЭВМ. Которые позволяют не только отсортировать материал по изучаемым признакам, но и выполнить расчеты показателей.

4 этап. Статистический анализ изучаемого явления, формулировка выводов — ответственный этап исследования, на котором проводится вычисление статистических показателей, дается их графическое изображение, изучается динамика, тенденции, устанавливаются связи между явлениями, даются прогнозы и т. д. Анализ предполагает интерпретацию полученных данных, оценку достоверности результатов исследования. В заключение делаются выводы.

5 этап. Литературная обработка и оформление полученных результатов — является заключительным, предполагает окончательное оформление результатов статистического исследования. Результаты могут быть оформлены в виде статьи, отчета, доклада, диссертации и др. Для каждого вида оформления существуют определенные требования, которые должны соблюдаться при литературной обработке результатов статистического исследования.

Результаты медико-статистического исследования внедряются в практику здравоохранения. Возможны различные варианты использования результатов исследования: ознакомление с результатами широкой аудитории медицинских и научных работников; подготовка инструктивно-методических документов; оформление рационализаторского предложения и другие.

По завершении статистического исследования разрабатываются рекомендации и управленческие решения, проводится внедрение результатов исследования в практику, оценивается эффективность.

В проведении статистического исследования важнейшим элементом является соблюдение строгой последовательности в осуществлении названных этапов.

Планирование клинического исследования

Структура научных исследований состоит из нескольких стандартных элементов, каждый из которых способствует тому, чтобы испытание дало ответ на поставленный исследователем клинический вопрос. Клинические исследования, посвященные оценке или разработке различных лекарственных препаратов (экспериментальные курсы лечения), как правило, дают ответы на вопросы об эффективности и безопасности этих лекарств.

А. Формирование гипотезы.

Любое исследование начинается с выдвижения гипотезы. *Гипотеза* — это сформулированное предположение о наличии или отсутствии взаимосвязи между какими-то событиями. Гипотезы позволяют проверить теории, на которых они основаны, определяют целесообразность проведения новых исследований, способствуют развитию различных научных областей. В статистике применяют понятие нулевой гипотезы — предположения, что полученные в исследовании различия случайны (обычно нулевая гипотеза формулируется как «обратное» тому, что требуется доказать. А хорошо знакомое всем p — это вероятность того, что сформулированная нулевая гипотеза верна, справедлива.

Таким образом, при $p < 0,05$ эта вероятность менее 5 %, что считается статистически значимым доказательством правильности утверждения, обратного нулевой гипотезе.

Б. Основные задачи экспериментального курса лечения.

Обычно в экспериментальном курсе лечения исследуется лекарственная форма (какую форму следует использовать — таблетки, капсулы, суспензию и т. д.); способ применения (перорально, ректально, трансдермально и т. д.); режим приема дозы; сроки лечения. Особенно важным представляется изучение типа дозирования — фиксированный, титрованный. При фиксированном типе дозирования в клинических испытаниях рандомизация (случайное распределение в группы) пациентов проводится в группы с разными дозировками препарата. В исследованиях с титрованным способом дозирования всем рандомизированным пациентам доза титруется либо до максимально переносимой, либо до оптимальной для достижения эффекта.

В. Цели и задачи исследования.

Основными целями клинических исследований, посвященных оценке терапии, являются определение терапевтического воздействия изучаемого препарата (или способа лечения) на конкретную группу пациентов по сравнению с контрольной группой, получение информации о побочных эффектах, решение вопросов об изменении качества жизни, зависимость

этих показателей от стоимости проводимого лечения. Таким образом, главными составляющими цели научного исследования являются:

1. Терапевтические эффекты данного вида терапии по сравнению с контрольным.

2. Побочные эффекты изучаемого лечения.

3. Критерии качества жизни пациентов и стоимость лечения.

Задачи исследования должны быть предельно конкретными, четкими. Все цели и задачи планируемого исследования должны определяться обязательно до его начала. Не следует в рамках одного клинического испытания ставить слишком много вопросов — это может явиться причиной недостоверности полученных ответов на большинство из них.

Г. Планирование эффективных и этических технологий контроля.

При проведении клинических исследований необходимо исключить возможные исходные различия включаемых пациентов, известный эффект плацебо, влияния внешних, несвязанных с лечением, условий и факторов.

Поэтому с целью оценки воздействия изучаемого лечения при проведении контролируемых клинических исследований используют четыре типа технологий контроля:

1. Контроль исходного состояния.

2. Плацебо-контроль.

3. Активный контроль.

4. Контроль по архивным данным (не относится к условиям проведения исследования).

Д. Мощность и объем выборки.

Особенностью планирования всех без исключения клинических исследований является то, что исследователь никогда не имеет в своем распоряжении всей популяции или генеральной совокупности. Обычно исследователи имеют дело с выборкой из этой совокупности. Любая выборка описывает генеральную совокупность с заведомыми погрешностями и искажениями, природу и степень выраженности которых необходимо понимать и учитывать при перенесении результатов исследований на общую популяцию. Поэтому уже на этапе планирования исследования особую важность приобретает решение вопроса об объеме выборки, достаточном для формирования статистически значимого заключения об эффективности проводимого лечения или о различиях в эффектах разных способов терапии.

Определение объема выборки (количества участников исследования) является одной из основных задач на этапе планирования клинических исследований. Часто на практике эта задача решается не строго, без использования соответствующих математических формул и правил статистики, а лишь основываясь на прошлом опыте проведения сходных исследований.

Следует помнить, что неэтичными считаются как исследования, имеющие чрезмерно большую численность включенных больных, так и

исследования со слишком малой выборкой, не позволяющие сформировать статистически значимое заключение об эффективности проводимого лечения, выявить различия или их отсутствие в эффекте изучаемой и контрольной терапии.

Е. Критерии отбора больных (включения/исключения).

Самым главным правилом на этапе отбора пациентов является формирование метода отбора, определение критериев включения и исключения из исследования еще до сбора каких-либо данных. Изменение критериев отбора после завершения набора пациентов или (что еще хуже) на этапе включения больных в исследование существенно искажают результаты и значительно снижают надежность испытания в целом.

Желательно, чтобы специалист, проводящий исследование, не знал в какую группу — основную или контрольную — попадет тот или иной пациент. Не должно быть различий и в сборе анамнеза у пациентов исследуемых групп. Определение критериев включения и исключения должно быть обоснованным, так как искусственное, необоснованное решение об исключении каких-либо больных может привести к обнаружению ложных связей или несуществующих различий. Одним из основных правил при определении критериев включения/исключения при возникающих сомнениях является сужение и ужесточение этих показателей. Четкие и однозначные результаты оценки эффективности и безопасности лечения, полученные в тщательно и правильно отобранных группах, представляют большую ценность, чем неоднозначные сведения, полученных на слишком разнообразных группах больных.

Ж. Методики контроля погрешности.

Грамотно спланированные клинические испытания позволяют исследователю с высокой долей уверенности судить о том, что данное воздействие лекарственного средства не было случайным, а выявленный эффект не явился результатом какого-то иного воздействия. Таким образом, исследователи делают выводы на основе вероятностного принципа, а не на основе 100 % доказательств, а статистическая обработка помогает оценить достоверность сделанных выводов. Поэтому никогда нельзя быть абсолютно уверенным в размере эффекта изучаемого лечения (причем, независимо от статистической мощности и грамотного планирования исследования) — ведь всегда возможна случайная ошибка.

2. ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ

Задача 1.

В городе N в 2016 году проживало 58320 человек, в течение года умерло 457 человек. Определить показатель смертности.

Задача 2.

В 2018 году число заболевших ОРВИ детей одной из средних школ составило 248 случаев, в том числе 117 мальчиков. Определить процентные доли случаев заболеваний мальчиков и девочек.

Задача 3.

В области N численность населения на начало 2012 года составляет 2 280 000 человек. В медицинских организациях области 20 670 больничных коек. Определить показатель обеспеченности населения больничными койками.

Задача 4.

В результате проведенного исследования по оценке влияния курения на смертность от рака легкого были получены следующие данные: смертность от рака легкого среди курящих составила 0,84 ‰ в год, смертность от рака легких среди некурящих — 0,05 ‰ в год, распространенность курения — 67 ‰, общая смертность от рака легкого — 0,49 ‰ в год. Рассчитать относительный риск, атрибутивный риск, добавочный популяционный риск и добавочную долю популяционного риска.

Задача 5.

Вероятность выздоровления при применении лекарственного препарата Артекс равна 75 ‰. Каково наиболее вероятное количество выздоровевших пациентов из 376 заболевших?

Задача 6.

Кумулятивная вероятность выживания после проведения пятилетнего курса лечения равна 0,7. Какое количество пациентов останется в живых, если рассматриваемая группа составляла пять лет назад 659 человек.

Задача 7.

Вероятность того, что больной останется жив спустя пять лет равна 0,75 (75 ‰), то вероятность того, что он за этот временной промежуток умрет, равна 0,25 (25 ‰). Каковы шансы у пациента выжить?

Задача 8.

Допустим, что некоторое лечение проходят и мужчины, и женщины. Вероятность того, что больной мужского пола останется жив спустя пять лет равна 0,75 (75 %); вероятность того, что он за этот временной промежуток умрёт, равна 0,25 (25 %). Аналогичные вероятности для женщин равны 0,7 и 0,3. Найти отношение шансов.

Задача 9.

Определить средний вес мальчиков 7 лет, относительные частоты, дисперсию, среднеквадратичное отклонение, коэффициент вариации (таблица 4).

Таблица 4 — Исходные данные для задачи

Вес, кг	15–18,9	19–22,9	23–26,9	27–30,9	31–34,9
Число мальчиков	16	27	32	16	9

Задача 10.

Изучалась длительность временной нетрудоспособности у больных с острыми респираторными заболеваниями. Было установлено, что в течение 5 дней болели 5 человек, 6 дней — 5; 7 дней — 5; 8 дней — 2; 9 дней — 5. Составить вариационный ряд и рассчитать основные его характеристики.

Задача 11.

Младенческая смертность в Республике Беларусь составила: в 1999 году — 11,4 ‰, в 2000 — 9,3‰, в 2001 — 9,1‰, в 2002 — 7,9 ‰. Представить данные графически, вычислить показатели динамического ряда: абсолютный прирост, темп прироста, темп роста, абсолютное значение 1 % прироста. Сделать выводы.

Задача 12.

Имеются данные о возрасте пациента, у которого впервые выявлено заболевание, и продолжительности лечения (таблица 5).

Таблица 5 — Исходные данные для задачи

Возраст, лет	44	16	36	26	45	21	38	24	18	26	20	26	27	25
Продолжительность лечения, дни	22	11	21	6	20	12	21	9	16	13	11	12	10	12

Определить количественную взаимосвязь между признаками. С помощью графического метода определить форму связи. Определить степень тесноты связи между рассматриваемыми признаками. Построить линию регрессии и на основе уравнения регрессии сделать прогноз продолжительности лечения пациента, возраст которого 30 лет. Объяснить смысл полученных параметров уравнения регрессии.

ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ

Пример 1.

В городе N в 2015 году проживало 46520 человек (среда), в течение года умерло 412 человек (явление).

Определить показатель смертности.

Решение:

$$\text{Смертность} = \frac{412}{46520} \times 1000 = 8,9 \text{ ‰} .$$

Пример 2.

В 2016 году число всех заболеваний детей в детском саду составило 102 случая энтеритов и дизентерии, в том числе 50 случаев энтеритов. Определить процентные доли случаев энтеритов и дизентерии.

Решение:

Если принять все случаи за 100 %, то:

$$\text{Доля энтеритов} = \frac{50}{102} \times 100 = 49 \text{ ‰} .$$

$$\text{Доля дизентерии} = \frac{52}{102} \times 100 = 51 \text{ ‰} .$$

Пример 3.

В области N численность населения на начало 2010 года составляет 1 235 000 человек. В медицинских организациях области 10 785 больничных коек. Определить показатель обеспеченности населения больничными койками.

Решение:

$$\text{Обеспеченность} = \frac{10785}{1235000} \times 10000 = 87 .$$

Таким образом, на каждые 10 000 населения развернуто 87 больничных коек.

Пример 4.

В результате проведенного исследования по оценке влияния курения на смертность от рака легкого были получены следующие данные: смертность от рака легкого среди курящих составила 0,96 ‰ в год, смертность от рака легких среди некурящих — 0,07 ‰ в год, распространенность курения — 56 %, общая смертность от рака легкого — 0,56 ‰ в год. Рассчитать относительный риск, атрибутивный риск, добавочный популяционный риск и добавочную долю популяционного риска.

Решение:

$$\text{Относительный риск} = \frac{0,96}{0,07} = 13,7.$$

$$\text{Атрибутивный риск} = 0,96 - 0,07 = 0,89.$$

$$\text{Добавочный популяционный риск} = 0,86 \times 0,56 = 0,5.$$

$$\text{Добавочная доля популяционного риска} = \frac{0,5}{0,56} \times 100 = 0,89 \times 100 = 89 \ %.$$

Пример 5.

Вероятность выздоровления при применении лекарственного препарата Аримидекс равна 70 %. Каково наиболее вероятное количество выздоровевших пациентов из 273 заболевших?

Решение:

Событием является «выздоровление больного», условием «больной принимает Аримидекс», степенью возможности — 70 % (грубо говоря, из 100 человек, принимающих Аримидекс, выздоравливают 70).

$$\text{Количество пациентов} = \frac{273}{100} \times 70 = 191,1 \approx 191 \text{ человек.}$$

Пример 6.

Кумулятивная вероятность выживания после проведения пятилетнего курса лечения равна 0,7. Какое количество пациентов останется в живых, если рассматриваемая группа составляла пять лет назад 784 человека.

Решение:

$$\text{Количество пациентов} = 784 \times 0,7 = 548,8 \approx 549 \text{ человек.}$$

Пример 7.

Вероятность того, что больной останется жив спустя пять лет равна 0,8 (80 %), то вероятность того, что он за этот временной промежуток умрёт, равна 0,2 (20 %). Каковы шансы у пациента выжить?

Решение:

$$\text{Шанс} = \frac{0,8}{0,2} = 4.$$

Таким образом, вероятность выздоровления в 4 раза больше вероятности смерти.

Пример 8.

Допустим, что некоторое лечение проходят и мужчины, и женщины. Вероятность того, что больной мужского пола останется жив спустя пять лет равна 0,6 (60 %); вероятность того, что он за этот временной промежуток умрёт, равна 0,4 (40 %). Аналогичные вероятности для женщин равны 0,8 и 0,2. Найти отношение шансов.

Решение:

$$\text{Отношение шансов} = \frac{0,6}{0,4} : \frac{0,8}{0,2} = 0,375.$$

Пример 9.

Определить средний рост юношей 17 лет, относительные частоты, дисперсию, среднеквадратичное отклонение, коэффициент вариации (таблицы 6 и 7).

Таблица 6 — Исходные данные для решения задачи

Рост, см	Число юношей
154–156	45
157–159	60
160–162	83
163–165	83
166–168	53
169–171	6

Решение:

Таблица 7 — Расчётная таблица

Интервал	Середина (t)	f _a	f _o	t • f _a	f _a • (t - \bar{x}) ²
154–156	155	45	0,14	6975	1911,90
157–159	158	60	0,18	9480	742,66
160–162	161	83	0,25	13363	22,29
163–165	164	83	0,25	13612	511,23
166–168	167	53	0,16	8851	1592,67
169–171	170	6	0,02	1020	431,65
Σ	—	330	1	53301	5212,39

$$\bar{x} = \frac{53301}{330} = 161,52$$

$$\sigma^2 = \frac{5212,39}{330} = 15,8$$

$$\sigma = \sqrt{15,8} = 3,97$$

$$C_v = \frac{3,97}{161,52} \cdot 100\% = 2,5\%$$

Так как $C_v < 10\%$, то разнообразие ряда считается слабым.

Пример 10.

Необходимо дать анализ динамики рождаемости в определенном районе (таблица 8).

Таблица 8 — Динамика рождаемости в регионе за 1996–2005 годы

Год	Рождаемость, %	Абсолютный прирост	Темп прироста, %	Темп роста, %	Абсолютное значение 1 % прироста
1996	9,4	—	—	—	—
1997	8,9	-0,5	-5,3	94,7	0,09
1998	9,2	0,3	3,4	103,4	0,09
1999	9,3	0,1	1,1	101,1	0,09
2000	9,4	0,1	1,1	101,1	0,09
2001	9,2	-0,2	-2,1	97,9	0,10
2002	8,9	-0,3	-3,3	96,7	0,09
2003	9,0	0,1	1,1	101,1	0,09
2004	9,1	0,1	1,1	101,1	0,09
2005	9,2	0,1	1,1	101,1	0,09

Решение:

Например, в 1997 году:

- абсолютный прирост = $8,9 - 9,4 = -0,5$;
- темп прироста = $(-0,5) \times 100/9,4 = -5,3 \%$;
- темп роста = $8,9 \times 100/9,4 = 94,7 \%$;
- абсолютное значение 1 % прироста = $0,5 / 5,3 = 0,09$.

Пример 11.

Имеются данные о возрасте пациента и количестве здоровых зубов (таблица 9).

Таблица 9 — Исходные данные для решения задачи

Возраст, лет	Количество зубов
20	24
22	26
25	25
27	21
30	18
31	15
32	17
35	18
38	14
40	12

1. Построить диаграмму рассеивания. Сделать предположения о наличии и силе корреляции. Объяснить.

2. Определить, количественную взаимосвязь между признаками. С помощью графического метода определить форму связи. Определить степень тесноты связи между рассматриваемыми признаками.

3. Построить линию регрессии и на основе уравнения регрессии сделать прогноз количества здоровых зубов для возраста 29 лет. Объяснить смысл полученных параметров уравнения регрессии.

Решение:

Таблица 10 — Расчетная таблица для коэффициента корреляции

	Возраст, лет(x)	Зубов (y)	xy	x ²	y ²
	20	24	480	400	576
	22	26	572	484	676
	25	25	625	625	625
	27	21	567	729	441
	30	18	540	900	324
	31	15	465	961	225
	32	17	544	1024	289
	35	18	630	1225	324
	38	14	532	1444	196
	40	12	480	1600	144
Сумма	300	190	5435	9392	3820

$$\bar{x} = \frac{300}{10} = 30$$

$$\bar{y} = \frac{190}{10} = 19$$

$$\overline{xy} = \frac{5435}{10} = 543,5$$

$$\overline{x^2} = \frac{9392}{10} = 939,2$$

$$r = \frac{10 \cdot 5435 - 300 \cdot 190}{\sqrt{(10 \cdot 9392 - 300^2)(10 \cdot 3820 - 190^2)}} = -0,92$$

Так как $r > 0,7$, то можно говорить о сильной отрицательной линейной корреляции.

Сравним абсолютную величину коэффициента корреляции с табличными значениями (Таблица «Критические значения коэффициентов корреляции r-Пирсона (r-Спирмена), Наследов А. Д., 2014). Так как полученное нами значение коэффициента корреляции по модулю (0,92) превышает табличное $r_{кр}(10; 0,001 \%) = 0,87$, мы можем с доверительной вероятностью 99,9 % констатировать наличие достоверной отрицательной линейной корреляции.

$$b = \frac{543,5 - 30 \cdot 19}{939,2 - 30^2} = -0,67$$

$$a = 19 + 0,67 \cdot 30 = 39,1$$

Уравнение регрессии (рисунок 4):

$$y = 39,1 - 0,67x$$

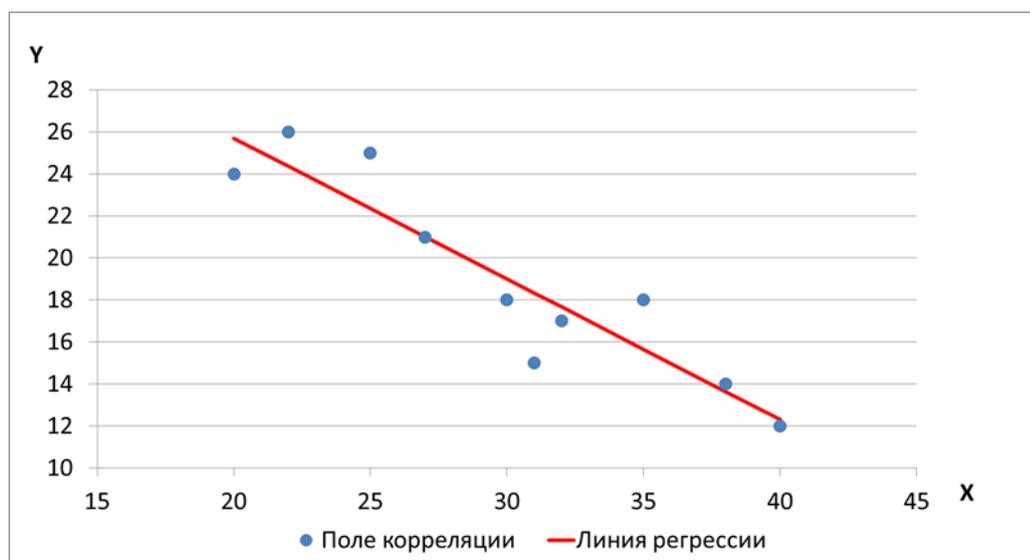


Рисунок 4 — Диаграмма рассеивания и линия регрессии

Прогноз для $x=29$ будет следующий $y = 39,1 - 0,67 \times 29 = 19,67$.

Следовательно, к 29 годам у человека прогнозируется 19–20 здоровых зубов.

Параметр $b < 0$ показывает, что зависимость отрицательная прямолинейная.

Параметр $b = -0,67$ показывает, что за каждый год количество здоровых зубов у человека сокращается на 0,67 (или за 3 года человек теряет примерно 2 зуба).

3. ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

1. Статистика, медицинская статистика, статистика здоровья, статистика здравоохранения.
2. Доказательная медицина. Критерии достоверности медицинской информации.
3. Основные источники статистической информации в здравоохранении (Белстат, Ежегодный статистический бюллетень, UNICEF Statistical Data, World Health Statistics, Global Health Observatory).
4. Признак и переменная. Количественные и качественные данные.
5. Генеральная совокупность и выборка, репрезентативность выборки.
6. Количественная и качественная однородность выборки.
7. Измерение. Шкала измерения, виды измерений.
8. Первичная и вторичная статистическая обработка данных.
9. Абсолютные и относительные статистические показатели, примеры.
10. Интенсивные и экстенсивные показатели.
11. Показатели соотношения и наглядности.
12. Средние величины, их виды.
13. Меры изменчивости, их виды.
14. Вероятность, шанс, отношение шансов.
15. Относительный, атрибутивный, добавочный популяционный риск.
16. Статистическое исследование и его этапы.
17. Планирование клинического исследования, его основные фазы.
18. Методы статистических исследований, когортное исследование.
19. Корреляция, её виды; коэффициент корреляции.
20. Линия регрессии, построение статистических прогнозов.

4. ТЕСТОВЫЙ КОНТРОЛЬ

Выберите один правильный вариант ответа

1. Предметом изучения медицинской статистики являются:

Варианты ответа:

- а) здоровье населения;
- б) влияние на здоровье населения различных факторов;
- в) деятельность лечебно-профилактических учреждений;
- г) воспроизводство населения;
- д) все перечисленное выше.

2. Характеризуют распределение явления на составные части, его внутреннюю структуру:

Варианты ответа:

- а) интенсивные показатели;
- б) экстенсивные показатели;
- в) абсолютные показатели;
- г) показатели наглядности;
- д) показатели соотношения.

3. Показатели обеспеченности населения врачами, больничными койками и т.п.:

Варианты ответа:

- а) интенсивные показатели;
- б) экстенсивные показатели;
- в) абсолютные показатели;
- г) показатели наглядности;
- д) показатели соотношения.

4. Структура населения по полу, возрасту, социальному положению и т.п.:

Варианты ответа:

- а) интенсивные показатели;
- б) экстенсивные показатели;
- в) абсолютные показатели;
- г) показатели наглядности;
- д) показатели соотношения.

5. Какой статистический показатель позволяет оценить тесноту взаимосвязи между стажем работы и уровнем заболеваемости работников?

Варианты ответа:

- а) среднее значение;

- б) дисперсия;
- в) коэффициент корреляции;
- г) коэффициент регрессии;
- д) коэффициент вариации.

6. Отношение каждого последующего уровня заболеваемости к предыдущему, выраженному в процентах:

Варианты ответа:

- а) абсолютный прирост;
- б) темп роста;
- в) темп прироста;
- г) абсолютный рост;
- д) средний прирост.

7. Разница между вероятностью развития заболевания в «экспонированной» и «неэкспонированной» группах:

Варианты ответа:

- а) атрибутивный риск;
- б) относительный риск.
- в) добавочный популяционный риск;
- г) доля атрибутивного риска;
- д) отношение шансов.

8. Позволяет оценить общий уровень признака для выборки, сравнить разные выборки или серии испытаний:

Варианты ответа:

- а) среднее значение;
- б) дисперсия;
- в) коэффициент корреляции;
- г) коэффициент регрессии;
- д) коэффициент вариации.

9. Ряд однородных сопоставимых величин, показывающих изменение изучаемого явления во времени (изменение уровня заболеваемости, смертности и т.п.):

Варианты ответа:

- а) интервальное распределение;
- б) вариационный ряд;
- в) частотное распределение;
- г) динамический ряд;
- д) нормальное распределение.

10. Проводят с целью установления абсолютных размеров явления (численности населения, общего количества родившихся или умерших):

Варианты ответа:

- а) сплошное статистическое наблюдение;
- б) текущее статистическое наблюдение;
- в) единовременное статистическое наблюдение;
- г) несплошное статистическое наблюдение;
- д) непосредственное статистическое наблюдение.

11. Способность выборки воспроизводить определенные характеристики генеральной совокупности в пределах допустимых погрешностей:

Варианты ответа:

- а) репрезентативность;
- б) достоверность;
- в) точность;
- г) однородность;
- д) воспроизводимость.

12. На какую величину в среднем изменится результативный признак при изменении факторного признака на единицу (ежегодный рост количества случаев профессиональных заболеваний):

Варианты ответа:

- а) среднее значение;
- б) дисперсия;
- в) коэффициент корреляции;
- г) коэффициент регрессии;
- д) коэффициент вариации.

ОТВЕТЫ К ТЕСТОВОМУ ЗАДАНИЮ

1	2	3	4	5	6	7	8	9	10	11	12
д	б	д	б	в	б	а	а	г	а	а	г

ЛИТЕРАТУРА

Основная

1. *Максимов, С. И.* Excel 2013 и SPSS 21 в решении задач прикладной статистики: учеб.-метод. пособие (с электронным приложением) / С. И. Максимов, Е. М. Зайцева. — 2-е изд., испр. — Минск: РИВШ, 2016. — 132 с.

2. *Максимов, С. И.* Технологии обработки данных исследований в IBM SPSS Statistucs: учеб.-метод. пособие / С. И. Максимов, Е. М. Зайцева. — Минск: РИВШ, 2016. — 100 с.

3. Использование MS Excel для анализа статистических данных: учеб. пособие / В. Р. Бараз, В. Ф. Пегашкин. — 2-е изд., перераб. и доп. — Нижний Тагил: НТИ (филиал) УрФУ, 2014. — 181 с.

4. Общественное здоровье и здравоохранение: учеб. пособие / Н. Н. Пилипцевич [и др.]; под ред. Н. Н. Пилипцевича. — Минск: Новое знание, 2015. — 784 с.

5. Общественное здоровье и здравоохранение: учебник: в 2 ч. / И. А. Наумов [и др.]; под ред. И. А. Наумова. — Минск: Высш. шк., 2013. — 351 с.

6. *Вальчук, Э. А.* Основы организационно-методической службы и статистического анализа в здравоохранении / Э. А. Вальчук, Н. И. Гулицкая, Ф. П. Царук. — Минск: 2003. — 381 с.

7. *Соболева, Л. Г.* Статистические величины. Использование их в медицине. Динамические ряды. Способы их выравнивания и анализа: учеб.-метод. пособие / Л. Г. Соболева, В. М. Дорофеев, Т. М. Шаршакова. — Гомель: ГГМУ, 2008. — 25 с.

8. *Шаршакова, Т. М.* Методы статистического анализа здоровья населения и деятельности организаций здравоохранения. Организация статистического исследования. Этапы статистического исследования. Графические изображения в статистике: учеб.-метод. пособие / Т. М. Шаршакова, Н. П. Петрова, Л. Г. Соболева — Гомель: ГГМУ, 2008. — 40 с.

Дополнительная

9. *Гржибовский, А. М.* Исследования типа «случай–контроль» в здравоохранении / А. М. Гржибовский, С. В. Иванов, М. А. Горбатова // Наука и здравоохранение. — 2015. — № 4. — С. 5–17.

10. *Гржибовский, А. М.* Когортные исследования в здравоохранении / А. М. Гржибовский, С. В. Иванов // Наука и здравоохранение. — 2015. — № 3. — С. 5–16.

3. Общая эпидемиология с основами доказательной медицины. Руководство к практическим занятиям: учеб. пособие / А. Ю. Бражников [и др.]; под ред. В. И. Покровского, Н. И. Брико. — 2-е изд., испр. и доп. — М.: ГЭОТАР–Медиа, 2012. — 496 с.

Учебное издание

Гапанович-Кайдалов Николай Владимирович
Шаршакова Тамара Михайловна

**МЕДИЦИНСКАЯ СТАТИСТИКА
И ОСНОВЫ ДОКАЗАТЕЛЬНОЙ МЕДИЦИНЫ**

**Учебно-методическое пособие
для студентов 4–6 курсов всех факультетов и магистрантов
учреждений высшего медицинского образования**

Редактор *Т. М. Кожемякина*
Компьютерная верстка *Ж. И. Цырыкова*

Подписано в печать 12.05.2020.
Формат 60×84^{1/16}. Бумага офсетная 80 г/м². Гарнитура «Гаймс».
Усл. печ. л. 2,56. Уч.-изд. л. 2,8. Тираж 90 экз. Заказ № 161.

Издатель и полиграфическое исполнение:
учреждение образования «Гомельский государственный медицинский университет».
Свидетельство о государственной регистрации издателя,
изготовителя, распространителя печатных изданий № 1/46 от 03.10.2013.
Ул. Ланге, 5, 246000, Гомель.