

ОСНОВЫ ПРАКТИЧЕСКОЙ БИО-МЕДИЦИНСКОЙ СТАТИСТИКИ

Часть 1

ПРИНЦИП СТАТИСТИКИ. ПОНЯТИЕ ВЕРОЯТНОСТИ. ВИДЫ ОШИБОК.
ПЛАНИРОВАНИЕ ИССЛЕДОВАНИЯ. НУЛЕВАЯ ГИПОТЕЗА. ПОНЯТИЕ ПОПУЛЯЦИИ
ИССЛЕДОВАНИЯ. РАСЧЕТ МОЩНОСТИ ИССЛЕДОВАНИЯ И РАЗМЕРА ВЫБОРКИ.
ОСНОВНЫЕ СТАТИСТИЧЕСКИЕ ПАКЕТЫ.

Современная
социология/экономик

Современная
философия

а

СТАТИСТИКА

Современная физика

Современная
биология/медицина

Примеры оценок (estimators)

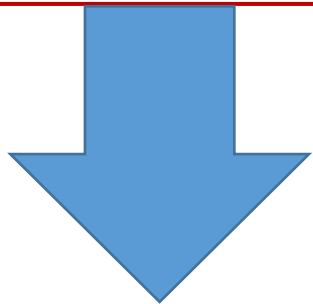
Оценка (estimator)	Оцениваемый показатель
Выборочное среднее, \bar{x}	Среднее ген.совок. μ
Выборочная доля, \hat{p}	Доля в ген.совок. p
Разность двух выборочных средних $\bar{x}_1 - \bar{x}_2$	Разность двух средних генеральной совокупности $\mu_1 - \mu_2$
Разность двух выборочных пропорций $\hat{p}_1 - \hat{p}_2$	Разность двух пропорций генеральной совокупности $p_1 - p_2$

Аналитическая (индуктивная) статистика (statistical inference) – процедура генерализации, в результате которой по данным из *выборки* делаются заключения о *генеральной совокупности*.

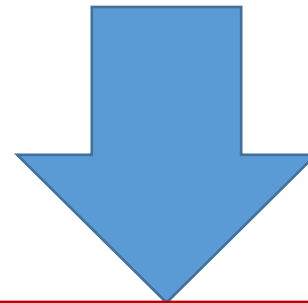
Основополагающим принципом при выборочной оценке (и, соответственно, для всей аналитической статистики) является *случайность выборки*

Вероятность – шанс что некоторое явление (необязательное)

произойдет
 $P(\text{выпадет туз из колоды карт}) = 4/52 = 0,0769.$



Фреквентистская
я (частотная
вероятность) –
предел к
которому
стремится шанс
события
произойти при
бесконечном
числе



Байесовская вероятность – вероятность
события, ЕСЛИ произошло другое,
взаимосвязанное с ним событие

$$P(A|B) = (P(B|A) * P(A)) / P(B)$$

Диагностический тест – чувствительность 90%, специфичность 99% (круто!)

Доля больных данным заболеванием – 0,001 (доля здоровых 0,999)

Найти вероятность того что данный человек здоров, если тест положителен!?

$P(B|A)$ – доля больных если предполагается изначально что человек
здоров – 0,01;
 $P(A)$ – доля здоровых – 0,999;

$P(B) = 0,999 \times 0,01 + 0,001 \times 0,9 = 0,01089;$

91,7%!!!

ПЛАНИРОВАНИЕ ИССЛЕДОВАНИЯ:

1. ИДЕЯ
2. СБОР И АНАЛИЗ ЛИТЕРАТУРЫ
3. ГИПОТЕЗА
4. ПЛАНИРОВАНИЕ ДИЗАЙНА ЭКСПЕРИМЕНТА
5. ВЫБОР СТАТИСТИЧЕСКОГО ТЕСТА
6. СБОР И АНАЛИЗ ЛИТЕРАТУРЫ
7. РАСЧЕТ МОЩНОСТИ

Тестирование гипотез

Гипотеза – утверждение, касающееся одной или нескольких популяций

1. **Исследовательская гипотеза** – предположение или вопрос, на разрешение которого направлено исследование (**PICO**)
2. **Статистическая гипотеза** – формулируется таким образом, что может быть *опровергнута* статистическими методами (inferential statistics)
 - H_0 – **нулевая гипотеза** (об отсутствии различий)
 - H_1 (H_A) – **альтернативная гипотеза** (комплементарна нулевой)

Нулевая и альтернативная гипотеза комплементарны, в сумме составляют все возможные вероятности, которые может принимать гипотетический показатель

Типы ошибок

Отбраковка верной нулевой гипотезы –
ошибка I рода, ее вероятность α

Невозможность отвергнуть ложную нулевую
гипотезу – **ошибка II рода**, ее вероятность β

Результат теста	На самом деле: H_0 верна	На самом деле: H_0 ошибочна
H_0 отвергнута	ошибка I рода, α	Верно (H_a)
H_0 отвергнуть не удалось	Верно (H_0) $1 - \alpha$	ошибка II рода, β

p

$1 - \beta$ –
мощность
исследования

Почему P-значения могут быть контр-интуитивны

1. Тестируемая гипотеза противоположна той гипотезе, которую мы хотим доказать (считаем верной)
2. Ученые и врачи могут считать «странным» рассчитывать теоретические распределения результатов экспериментов, которые никогда не будут проведены
3. «Логика» подхода интуитивно противоречит тому, что происходит на самом деле
 - Мы делаем выборку и хотим сделать заключения о популяции
 - Расчеты p-значений начинаются с предположений относительно популяции (нулевая гипотеза) и определяют вероятность обнаружить такие же или большие различия

Общий подход к тестированию гипотез

1. Сформулируйте научную гипотезу для исследования
2. Сформулируйте релевантную статистическую гипотезу (H_a)
3. Для статистической гипотезы подберите дополняющую нулевую гипотезу (H_0)
4. Рассчитайте вероятность получения различных (экспериментальных) данных
5. Определите долю **ВОЗМОЖНЫХ** результатов, для которых статистики теста будут **такой же или большей** величины
6. Сравните с заранее заданным пороговым значением

Теория значимости различий Фишера

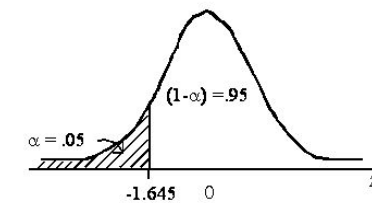
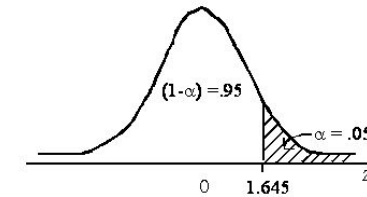
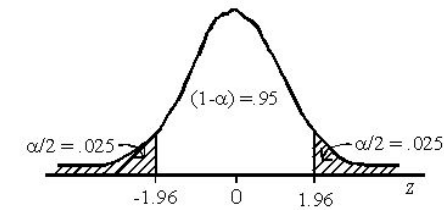
1. Определить нулевую статистическую гипотезу («нулевая» \neq «гипотеза об отсутствии всего»)
2. Определить точное значение p (например, $p=0,051$ или $p=0,049$). Ничего не говорить о принятии или опровержении гипотезы
3. Использовать данную процедуру только если об исследуемой проблеме известно очень мало

Теория планирования эксперимента Неймана-Пирсона

1. Установить две статистические гипотезы, H_1 и H_2 , а также принять решения относительно α , β , размера выборки до начала эксперимента, основываясь на субъективных представлениях о затратах и пользе. Эти решения определяют область [результатов], в которой каждая гипотеза будет отвергнута (rejection region, область отвержения)
2. Если результаты попадают в область отвержения для H_1 , принять гипотезу H_2 , в противном случае принять H_1 . «Принять» – не значит поверить (скорее, **действовать, как если бы она была верна**).

Одно- и двусторонние гипотезы

	H_0	H_a	Характер распределения и положение критической области
=	$\mu=50$	$\mu \neq 50$	Два «хвоста», две критических области
\leq	$\mu \leq 50$	$\mu > 50$	Один «хвост», критическая область справа
\geq	$\mu \geq 50$	$\mu < 50$	Один «хвост», критическая область слева



Тестирование гипотез

Важное замечание: ни тестирование гипотезы, ни статистический анализ не доказывают гипотезу. Заключение делается относительно того, ***опровергают ли полученные данные гипотезу***

ПЛАНИРОВАНИЕ ИССЛЕДОВАНИЯ:

1. ИДЕЯ
2. СБОР И АНАЛИЗ ЛИТЕРАТУРЫ
3. ГИПОТЕЗА
4. ПЛАНИРОВАНИЕ ДИЗАЙНА ЭКСПЕРИМЕНТА
5. ВЫБОР СТАТИСТИЧЕСКОГО ТЕСТА
6. РАСЧЕТ МОЩНОСТИ

Правильная формулировка клинического вопроса

Структура клинического вопроса: PICO(T)

Population/Patients – популяция/пациенты

Intervention - вмешательство

Comparison – группа сравнения

Outcome - исход

Time - время

Популяция и выборка

• ПОПУЛЯЦИЯ ИССЛЕДОВАНИЯ – генеральная совокупность, к которой относится выборка исследования и на которую можно генерализовать его результаты

- **Популяция** – это большая группа людей, проживающих в определенном географическом регионе (например, в Московской области) или обладающих некоторым признаком (например, старше 65 лет)
- Популяция может состоять из пациентов, госпитализированных в определенную клинику или из пациентов с определенным заболеванием
- **Выборка** – часть популяции, полученная путем отбора

ТИПЫ ИССЛЕДОВАНИЙ



Обсервационные

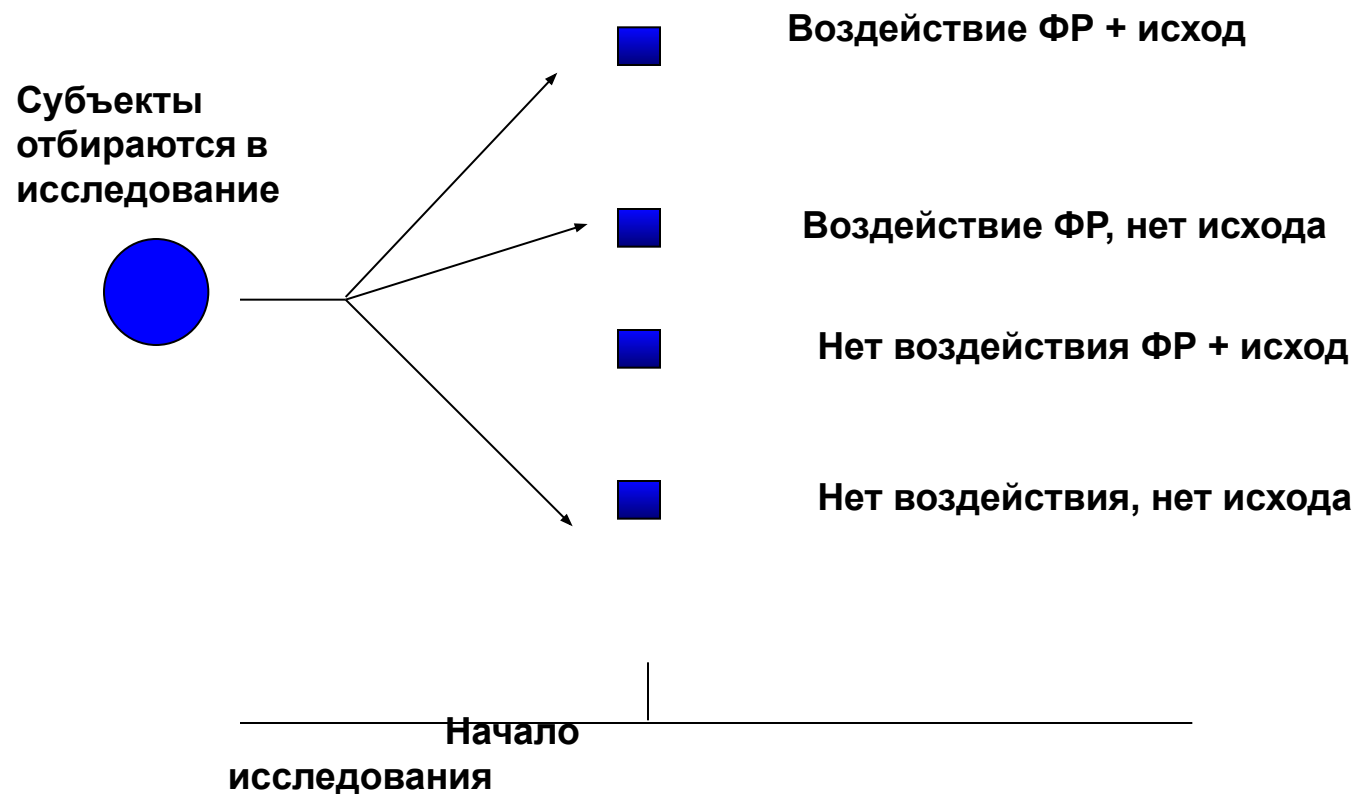
- *Поперечные (кросс-секционные)*
- *Случай-контроль*
- *Проспективное когортное*
 - *Ретроспективное когортное*

Экспериментальные

- *Двойное слепое рандомизированное*

Все факторы кроме одного фиксированы!

Кросс-секционные исследования



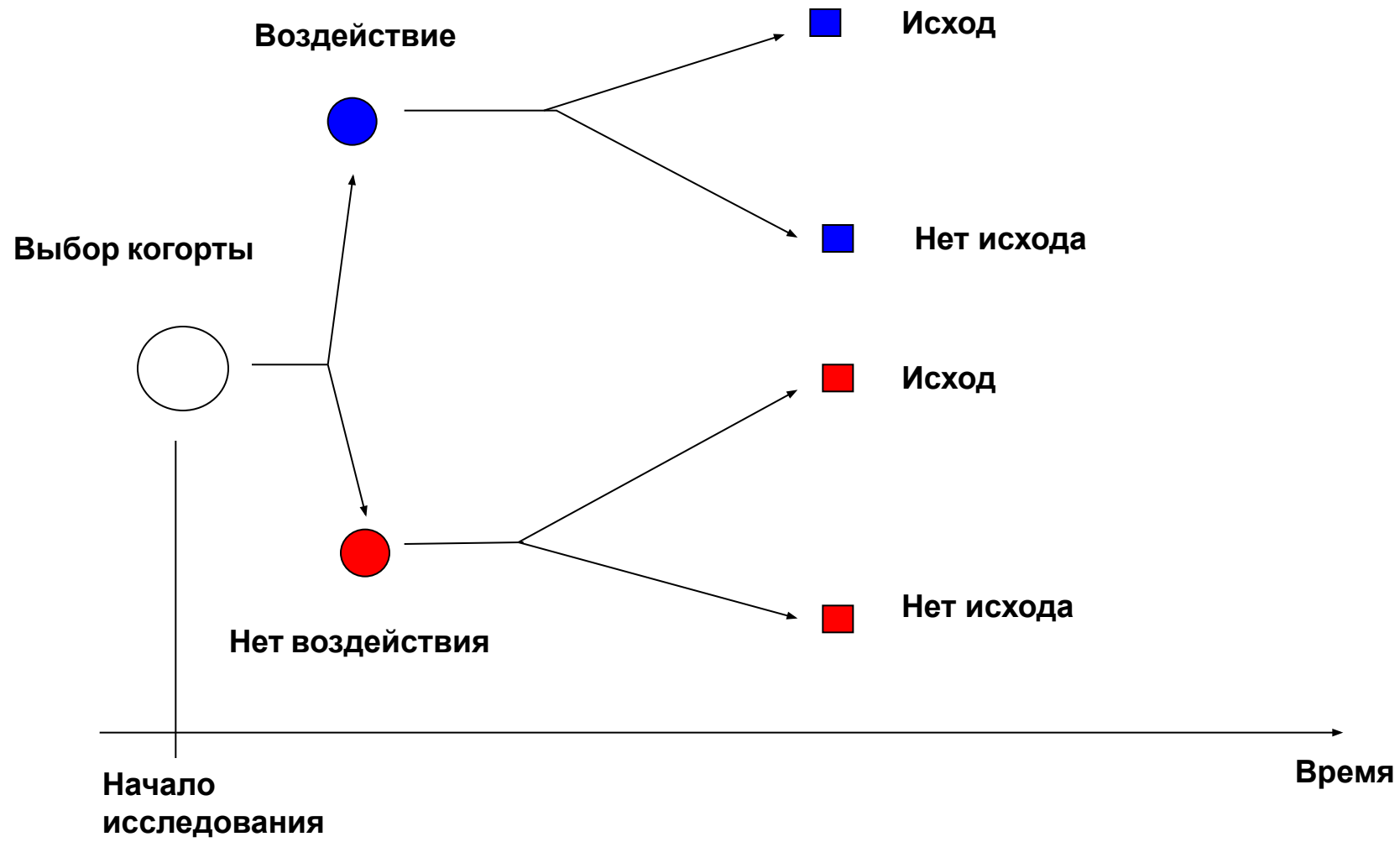
Кросс-секционные исследования (исследования распространенности)

- Анализ группы субъектов в один момент времени
- Позволяют описать заболевание и его важность (распространенность)
- Определяют потребность в диагностике/лечении
- Могут подразделяться на:
 - Описательные (результат - частоты)
 - Аналитические (результат - odds ratio)

Кросс-секционные исследования (исследования распространенности)

- Преимущества:
 - Позволяют узнать распространенность болезни в группе
 - Полезны для оценки диагностических процедур
 - Полезны для изучения распространенных факторов риска
 - Полезны для изучения распространенных исходов
- Недостатки:
 - Популяции субъектов как правило, не хотят принимать участие в исследованиях
 - Выборка получается нерепрезентативной
 - Бесплезно искать причины исходов

Когортные исследования



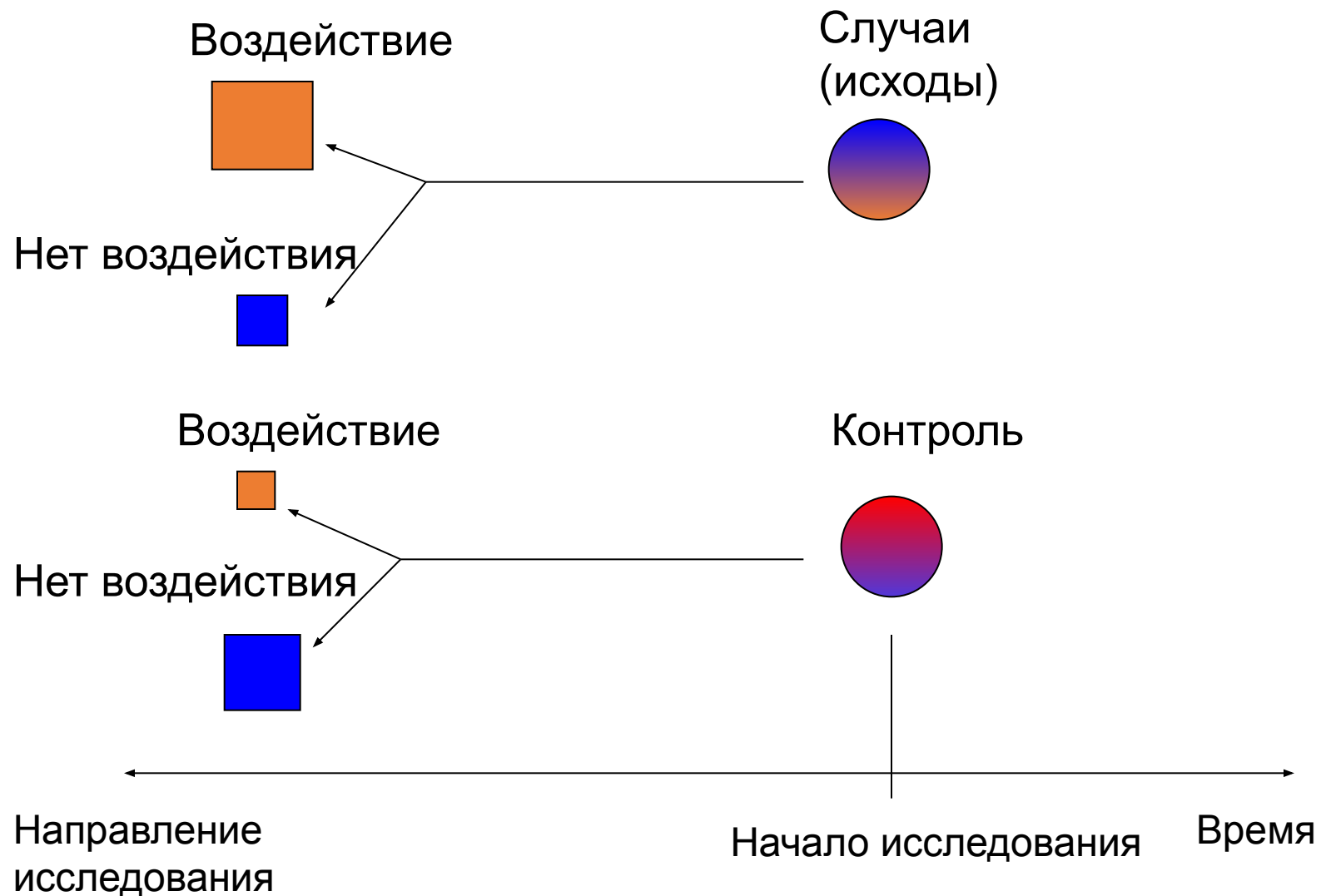
Когортные исследования

- Субъекты отбираются по принципу **отсутствия заболевания** (исхода) и классифицируются по наличию/отсутствию факторов риска
- Проводится проспективное наблюдение, направленное на **выявление исхода**
- Когортное исследование может быть **проспективным** и **ретроспективным**

Когортные исследования

- Преимущества:
 - Подходят для выявления причин заболевания/исхода
 - Подходят для выявления течения заболевания
 - Полезны, когда мы изучаем два или более исхода одновременно
- Недостатки:
 - Занимают длительное время
 - Дороги
 - Субъекты исследования склонны пропадать
 - Нерелевантны для изучения редких исходов

Исследования «случай – контроль»

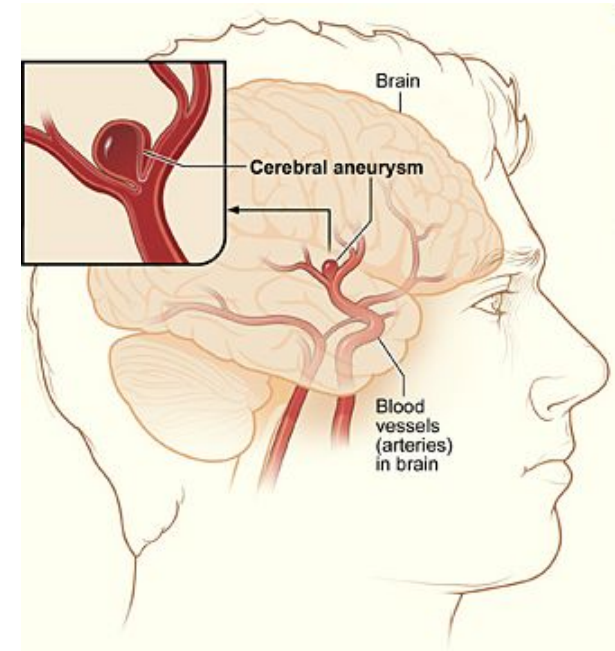


Исследования «случай-контроль»

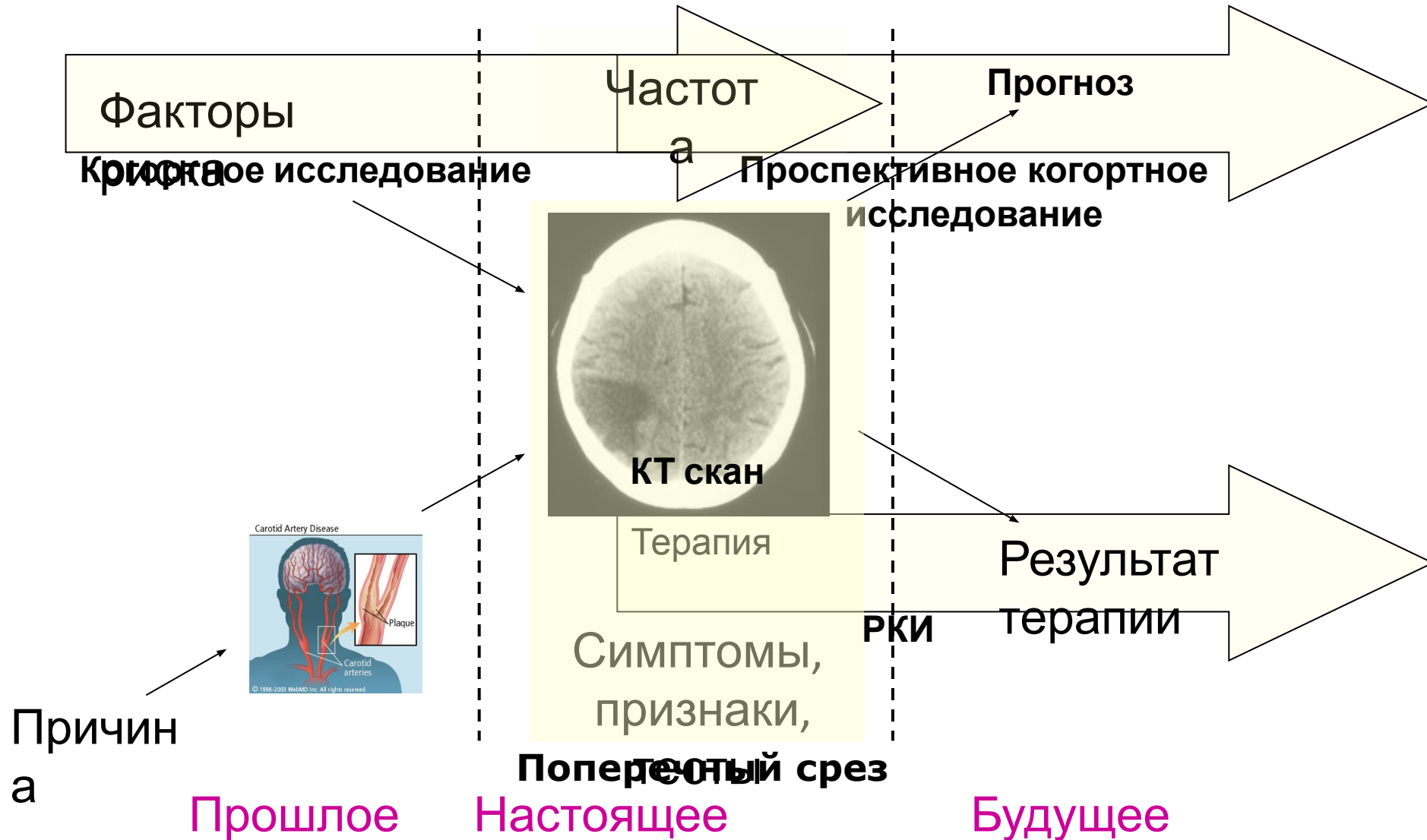
- Преимущества:
 - Подходят для изучения редких исходов
 - Адекватны для изучения исходов с долгим периодом развития
 - Требуют мало ресурсов на проведение
 - Нет необходимости ждать наступления исхода
- Недостатки:
 - Множество источников систематической ошибки
 - Зависят от количества и качества регистров пациентов
 - Контрольная группа должна быть адекватно подобрана, поскольку представляет собой популяцию без исходов

Формулировка клинического вопроса

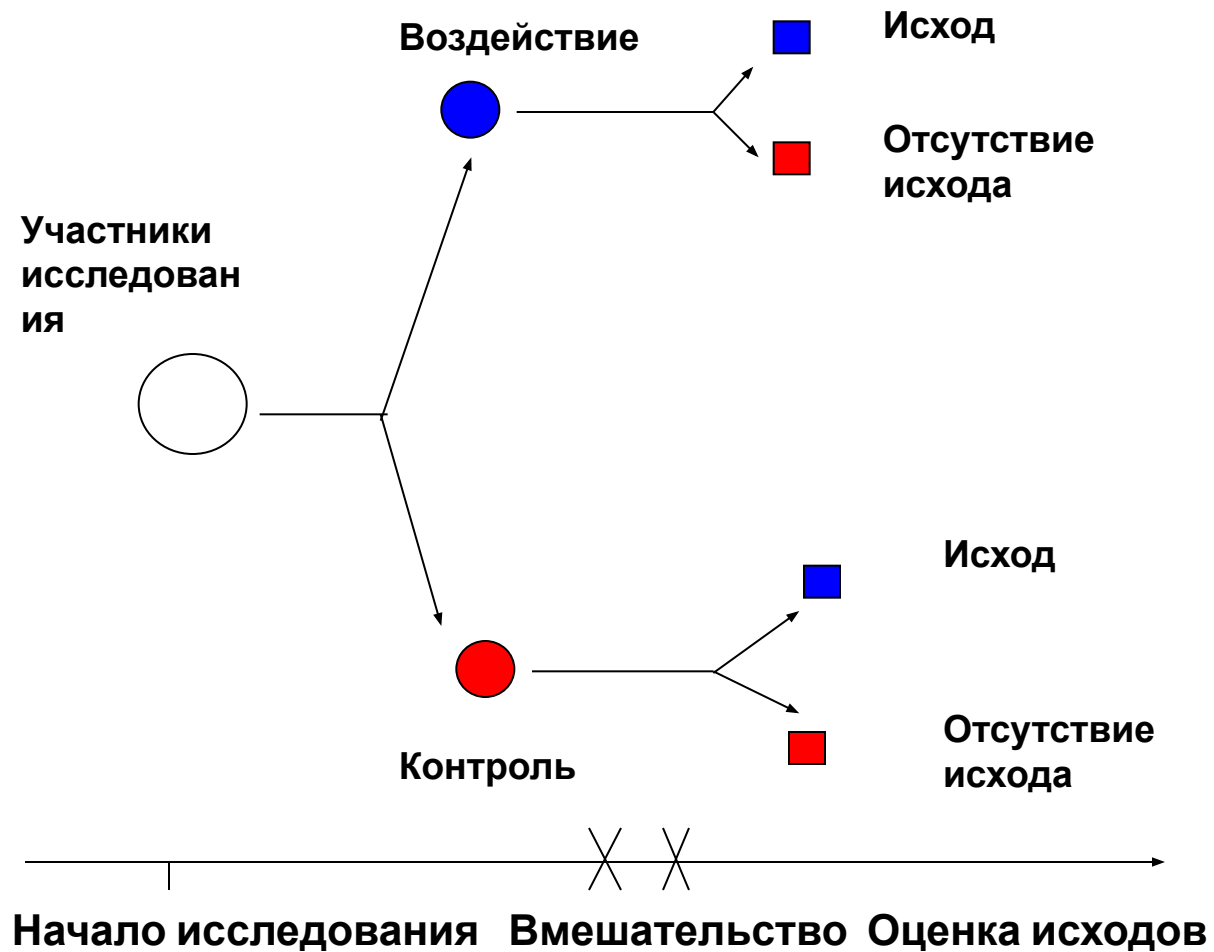
35-летний мужчина, брат которого недавно умер от разрыва аневризмы сосудов головного мозга, беспокоится, что у него также может быть аневризма и какова вероятность того, что она разорвется?



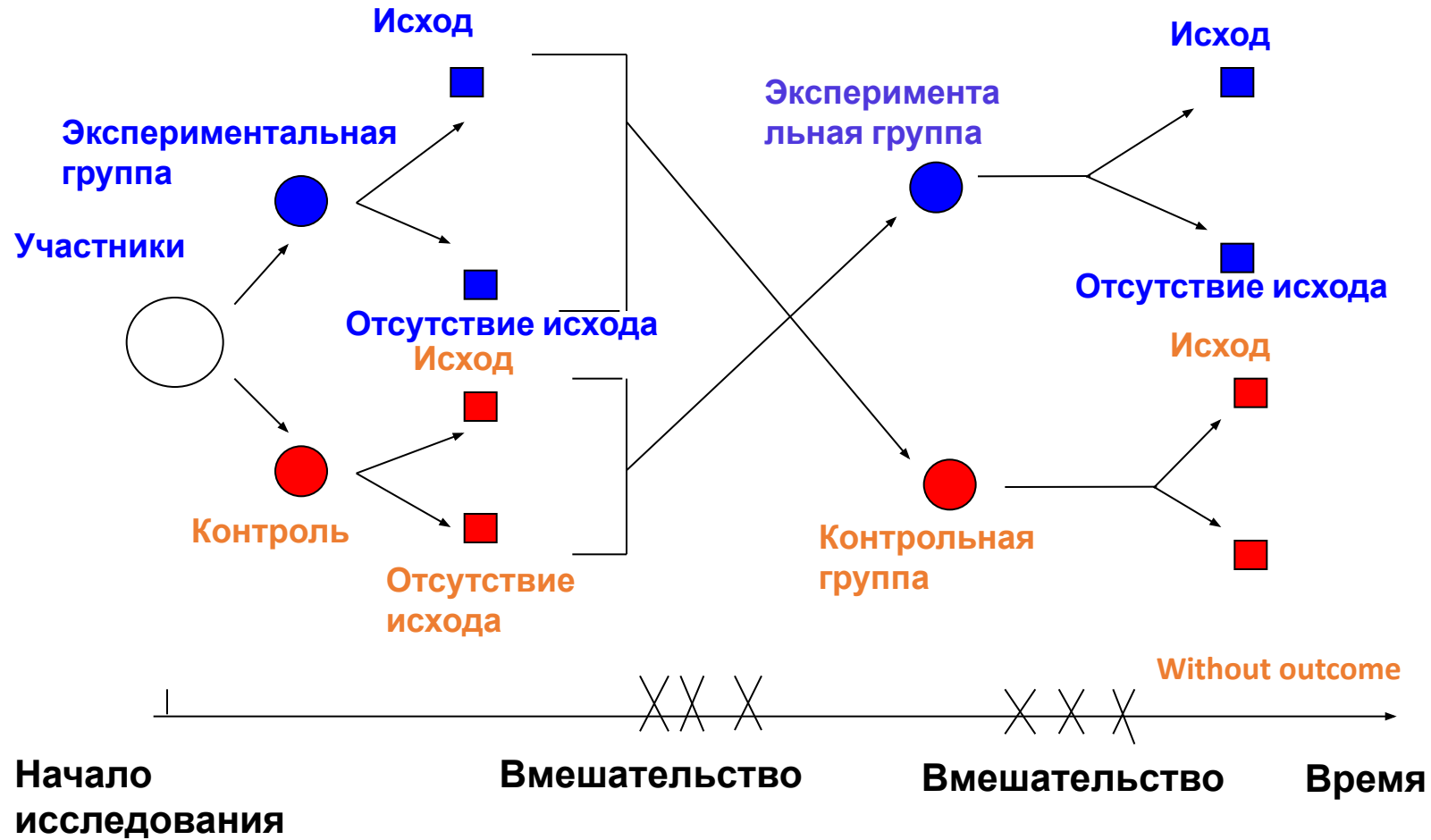
Типы исследований: случай с разрывом аневризмы



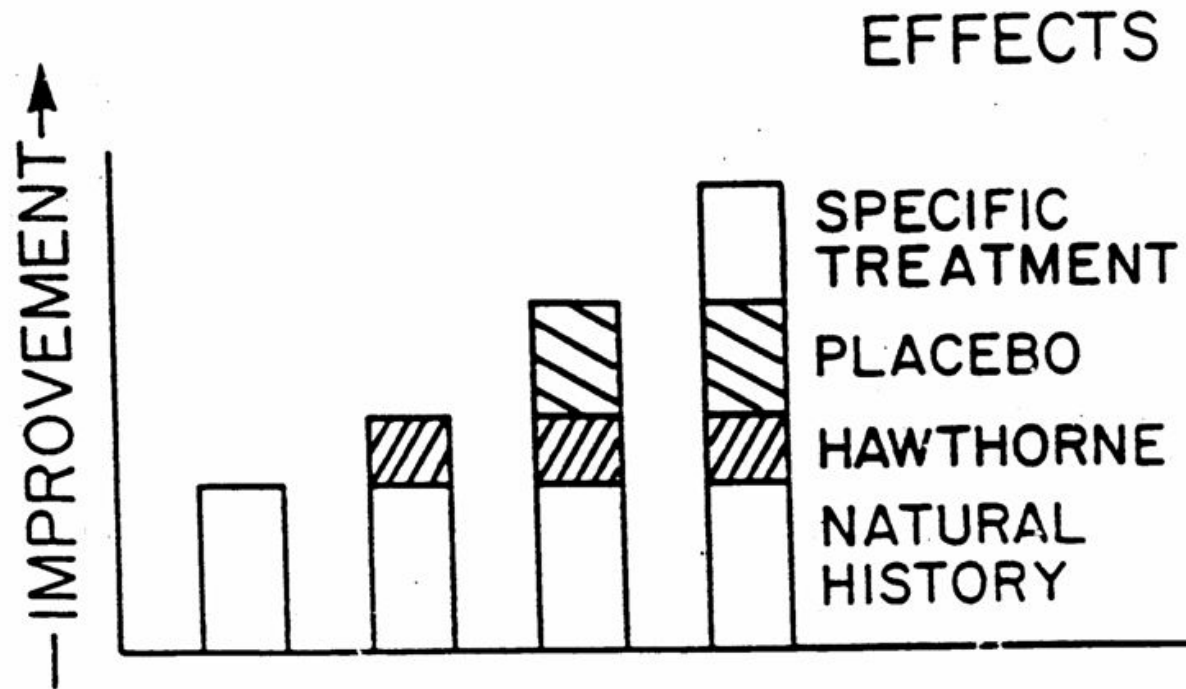
Экспериментальные исследования в параллельных группах



Экспериментальные исследования перекрестного дизайна



Общий эффект лечения является совокупностью эффектов от спонтанного улучшения, неспецифических реакций и эффектов специфической терапии



Хоторнский эффект (*Hawthorne effect*) - Участники эксперимента действуют иначе, более усердно, чем обычно, только благодаря осознанию того, что они причастны к эксперименту.

Вопросы при выборе показателя:

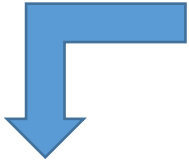
- Природа показателя
- Процесс измерения
- Характеристики измерения
(надежность/валидность)
- Выбор измерений
- Факторы, влияющие на измерения
- Когда и как часто измерение проводится в ходе исследования?

О выборе зависимой переменной!

Характеристики измерения:

- Некий объект или явление, которое изменяется наблюдаемым и количественно измеримым образом
 - Измерение = истинное значение + ошибка
 - Ошибка = случайная + систематическая

ОШИБКИ ИССЛЕДОВАНИЯ



СИСТЕМАТИЧЕСКИЕ

- СВЯЗАННЫЕ С ОТБОРОМ (когда группы в выборке отличаются по более чем одному параметру ИЛИ выборка подобрана так, что группы БУДУТ отличаться)
- СВЯЗАННЫЕ С ИЗМЕРЕНИЕМ (разные методы измерения в разных группах)
- СВЯЗАННЫЕ С ВЗАИМОДЕЙСТВИЕМ ФАКТОРОВ (CONFOUNDING, как правило не в экспериментальных исследованиях, накладывается на ошибку отбора)

СЛУЧАЙНЫЕ

- ОШИБКИ ИЗМЕРЕНИЯ (недостаточная выборка, ошибки собственно измерения (прибор сбоит))
- ОШИБКИ РЕГИСТРАЦИИ (записано неправильно)
- ОШИБКИ ПЕРЕНОСА/ОБРАБОТКИ (компьютер)

Подробнее

1. Систематическая ошибка отбора
2. Систематическая ошибка участия
3. Систематическая ошибка выбывания
4. Систематическая ошибка оценки эффективности
5. Систематическая ошибка репортирования
6. Систематическая ошибка, обусловленная вмешивающимися факторами

Способы контроля систематических ошибок

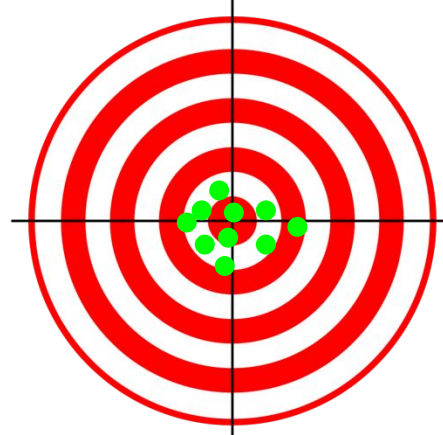
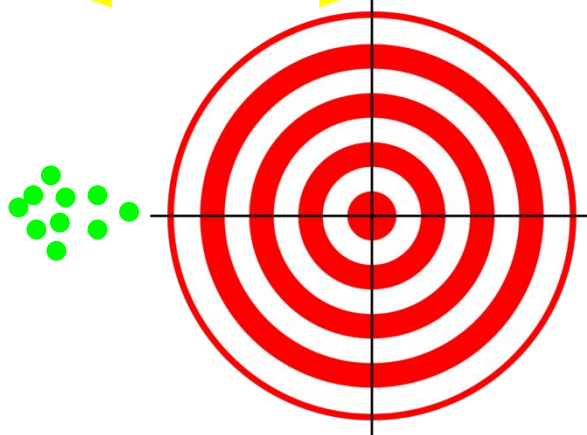
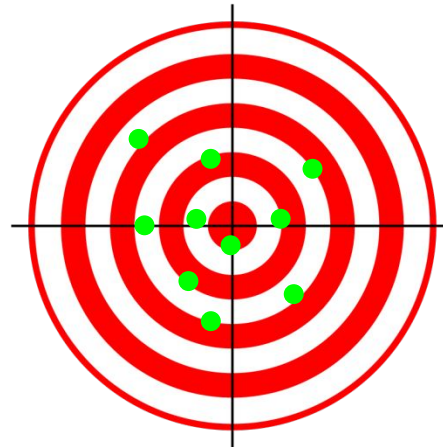
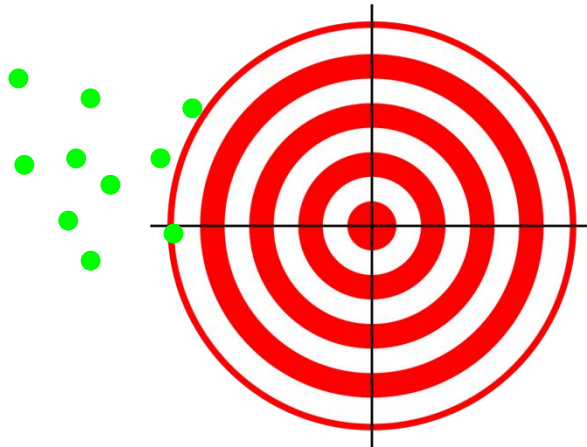
Систематическая ошибка	Способ контроля
Отбора (<i>selection bias</i>)	<ul style="list-style-type: none">• Рандомизация• Маскировка рандомизационной последовательности
Участия (<i>performance bias</i>)	<ul style="list-style-type: none">• Ослепление
Выбывания (<i>attrition bias</i>)	<ul style="list-style-type: none">• Полнота показателей эффективности• Ослепление
Оценки эффективности (<i>detection bias</i>)	<ul style="list-style-type: none">• Ослепление
Репортирования (<i>reporting bias</i>)	<ul style="list-style-type: none">• Контроль селективного репортирования• Выполнение протокола и плана статистического анализа

Характеристики измерения

Смещение (Bias)

Высокое (high bias)

Низкое (low bias)



Дисперсия (Variance)

Высокая (high variance)

Низкая (low variance)

Выбор конечных точек на

практике:

1. Формулировка цели исследования
 - Научная гипотеза, с указанием терапевтической области и популяции исследования (это не шутка!)
2. Поиск в рекомендациях регуляторных органов
 - ФГБУ «Научный центр экспертизы средств медицинского применения», Руководство по экспертизе лекарственных средств (вплоть до рекомендованных методов оценки)
 - Guidance for Industry: Clinical Trial Endpoints (FDA, EMA)
 - Прецеденты одобрения конечных точек
3. Поиск в результатах проведенных исследований
 - Результаты мета-анализа (сравнение с другими препаратами)
 - Результаты контролируемых исследований
 - Конечные точки одобренных (не завершившихся) исследований
 - Конечные точки в научных публикациях (не всегда РКИ)

На практике:

4. Конечный результат (иногда) требуется согласовать с opinion-leader`ом (разные школы)
 - При этом нужно учесть, что обычно OL в РКИ ни за что не отвечает и его рекомендации тоже требуют проверки (Эр)
5. Оценить имеющиеся данные, необходимые для планирования
 - Как минимум, ожидаемую величину эффекта (Ал)
 - Как минимум, оценку дисперсии показателя (Ал)
 - Часто – клинически значимые различия (ФГБУ - возвраты)
 - Часто – эффективность конкурирующих препаратов
 - Обязательно – «общую оценку эффективности» (Эр)
6. Оценить несколько вариантов развития событий
 - Оптимистический и пессимистический (как эффект, так и само исследование, удобство заполнения шкал)
 - Поиграть в «бэклог» (ок, мы провалили исследование, почему?)
 - Обдумать варианты дизайна (Повторные измерения? Перекрестный? Промежуточный анализ?)

При прочих равных:

- Исследования с повторными наблюдениями позволяют существенно увеличить мощность исследования
- Интервальная шкала (надежная, валидная, с достаточной для планирования информацией) позволяет существенно увеличить мощность исследования
- Время до наступления события хорошо работает только в крупномасштабных длительных исследованиях (азбука, но: в трех исследованиях за последний год Спонсор всеми силами настаивал на конечной точке)
- Качество жизни (SF-36 v2), несмотря на широкое применение, не лишено ряда спорных моментов (что произойдет со шкалой физического здоровья, если у пациента улучшилось психическое, и наоборот?)

ПЛАНИРОВАНИЕ ИССЛЕДОВАНИЯ:

1. ИДЕЯ
2. СБОР И АНАЛИЗ ЛИТЕРАТУРЫ
3. ГИПОТЕЗА
4. ПЛАНИРОВАНИЕ ДИЗАЙНА ЭКСПЕРИМЕНТА
5. ВЫБОР СТАТИСТИЧЕСКОГО ТЕСТА
6. РАСЧЕТ МОЩНОСТИ

Выбор статистического теста

Цель	Интервальная	Ранг, оценка	Биномиальное	Выживаемость
Описать 1 группу	Среднее, ст. отклонение	Медиана, IQR	Пропорция	Кривая выживаемости Каплана-Мейера
Сравнить 1 группу с гипотетическим значением	t-test для одной выборки	Тест Уилкоксона	Тест хи-квадрат	
Сравнить не связанные совокупности	t-тест для не связанных совокупностей	Тест Манн-Уитни	Точный критерий Фишера, хи-квадрат	Лог-ранговый тест или тест Мантеля-Ханзеля
Сравнить 2 связанные совокупности	t-тест для связанных совокупностей	Тест Вилкоксона	Тест Мак-Неймера	Условная регрессия пропорционального риска*

Number of Dependent Variables	Nature of Independent Variables	<u>Nature of Dependent Variable(s)</u>	Test(s)
1	0 IVs (1 population)	interval & normal	one-sample t-test
		ordinal or interval	one-sample median
		categorical (2 categories)	binomial test
		categorical	Chi-square goodness-of-fit
	1 IV with 2 levels (independent groups)	interval & normal	2 independent sample t-test
		ordinal or interval	Wilcoxon-Mann Whitney test
		categorical	Chi-square test
			Fisher's exact test
	1 IV with 2 or more levels (independent groups)	interval & normal	one-way ANOVA
		ordinal or interval	Kruskal Wallis
categorical		Chi-square test	
1 IV with 2 levels (dependent/matched groups)	interval & normal	paired t-test	
	ordinal or interval	Wilcoxon signed ranks test	
	categorical	McNemar	
1 IV with 2 or more levels (dependent/matched groups)	interval & normal	one-way repeated measures ANOVA	
	ordinal or interval	Friedman test	
	categorical	repeated measures logistic regression	
2 or more IVs (independent groups)	interval & normal	factorial ANOVA	
	ordinal or interval	ordered logistic regression	
	categorical	factorial logistic regression	
			correlation

1 interval IV	interval & normal	correlation	
		simple linear regression	
	ordinal or interval	non-parametric correlation	
	categorical	simple logistic regression	
1 or more interval IVs and/or 1 or more categorical IVs	interval & normal	multiple regression	
		analysis of covariance	
	categorical	multiple logistic regression	
		discriminant analysis	
2 or more	1 IV with 2 or more levels (independent groups)	interval & normal	one-way MANOVA
2 or more	2 or more	interval & normal	multivariate multiple linear regression
2 sets of 2 or more	0	interval & normal	canonical correlation
2 or more	0	interval & normal	factor analysis
Number of Dependent Variables	Nature of Independent Variables	<u>Nature of Dependent Variable(s)</u>	Test(s)

Планирование мощности исследования

- Концепция планирования мощности подразумевает определение необходимого размера выборки, чтобы обнаружить клинически/научно значимое различие для фиксированного уровня ошибки I рода

Планирование мощности исследования

- Формулирование статистических гипотез и методов статистического анализа, направленных на подтверждение этих гипотез;
- Определение максимально допустимых значений ошибок I и II рода (вероятность ошибки I рода и мощности исследования);
- Определение параметров распределения переменной отклика, включая показатели среднего и дисперсии в условиях отсутствия воздействия;
- Определение величины значимого клинического эффекта (детализация альтернативной гипотезы);
- Определение размера выборки клинического исследования;

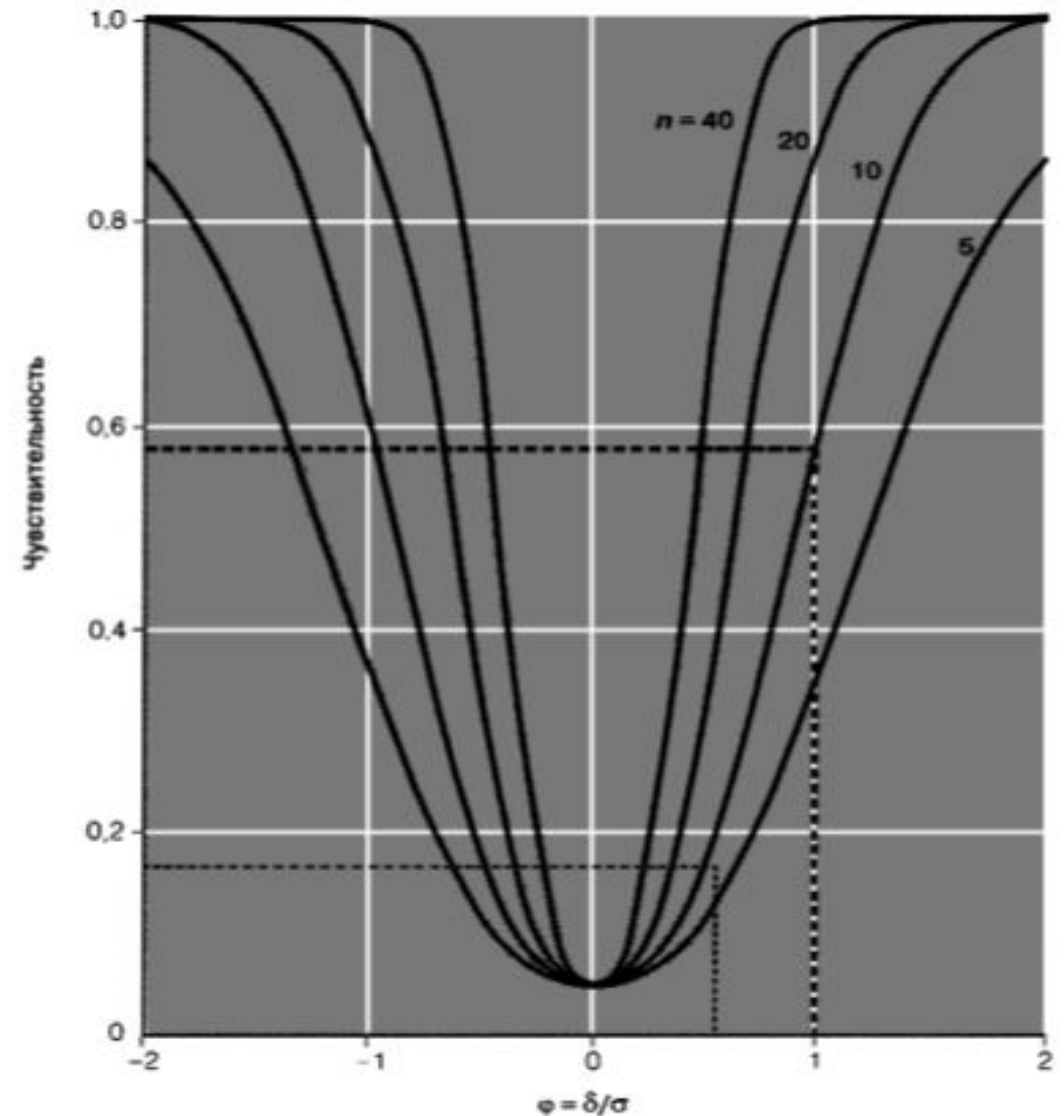
$\phi = \delta/\sigma$; где

ϕ – параметр нецентральности, σ – стандартное отклонение,
 δ – предполагаемая разница средни

Какова мощность исследования (чувствительность статистического метода), если разница средних – 20, а стандартное отклонение 15? В группах у нас по 10 человек.

Сколько нужно набрать крыс в группу для получения мощности исследования в 80%, если предполагаемая разница средних 0,3, а стандартное отклонение 2,0?

КРИТЕРИЙ СТЬЮДЕНТА



ДИСПЕРСИОННЫЙ АНАЛИЗ

$$\varphi = \frac{\delta}{\sigma} \sqrt{\frac{n}{2k}},$$

где, обозначения те же, кроме:

δ – МИНИМАЛЬНАЯ РАЗНИЦА МЕЖДУ ЛЮБЫМИ ДВУМЯ ГРУППАМИ;

k – число групп;

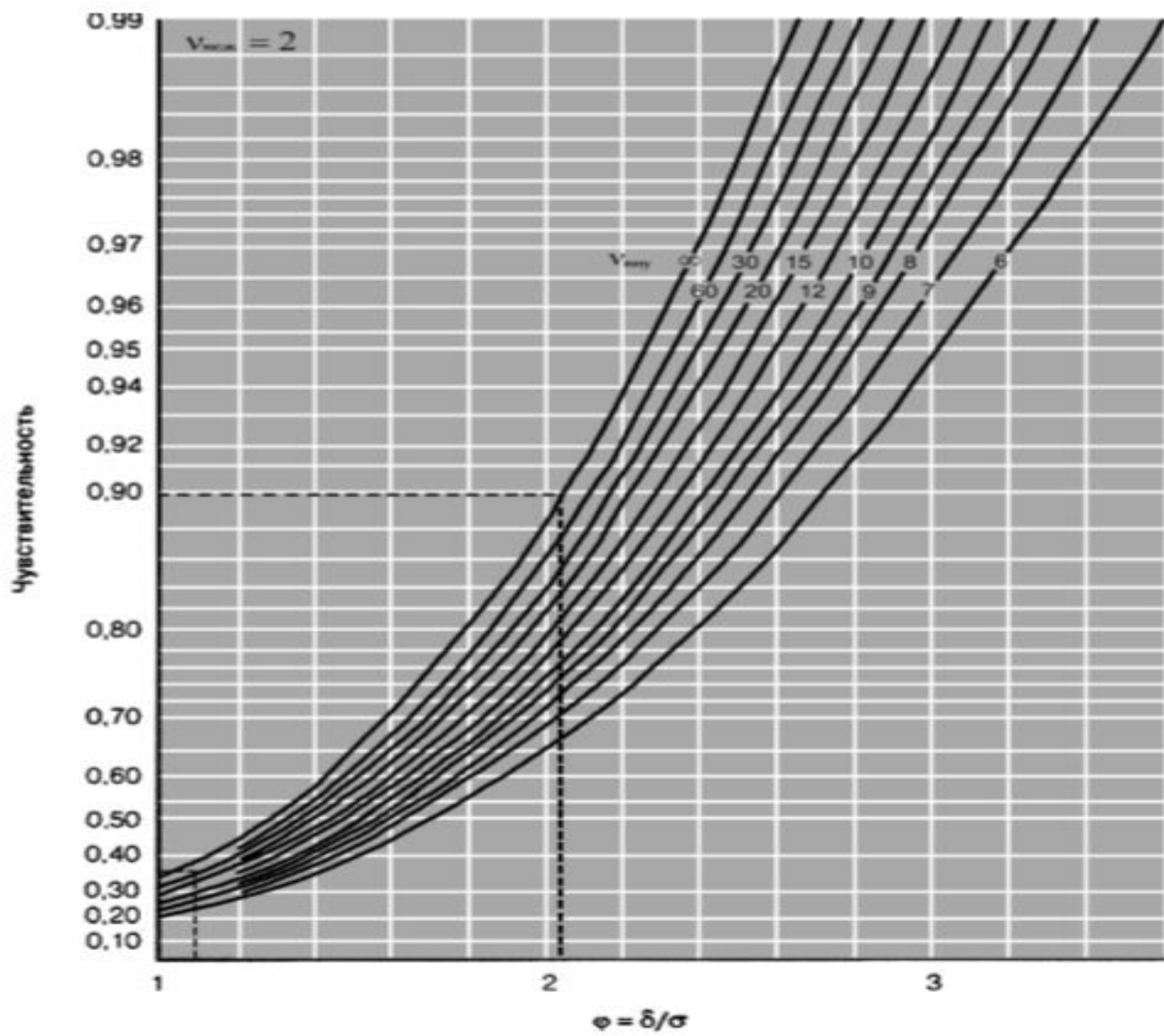
n – численность меньшей из групп(!).

v (меж) = $k - 1$, где,

v (меж) – число межгрупповых степеней свободы;

k – как раньше.

v (вну) = $n - k$.



Планирование объема выборки

Sample Size Calculations in Clinical Research

Second Edition

Shein-Chung Chow

Duke University School of Medicine
Duham, North Carolina, U.S.A.

Jun Shao

University of Wisconsin
Madison, U.S.A.

Hansheng Wang

Peking University
Beijing, China

Chapman & Hall/CRC

Taylor & Francis Group
Boca Raton London New York

2008 by Taylor & Francis Group, LLC

1. Excel

- Создание таблиц данных
- Хранение данных
- Фильтрация/сортировка
- Простейшие статистические анализы
- Неплохой но не всегда удобный графический модуль

2. Statistica

- То же, но больше опций для анализа
- Работа с переменными неудобна

3. SPSS

- Возможности чуть больше чем у Statistica
- Удобная работа с переменными

Возможна работа с базами данных

4. SAS

- Может все
- Золотой стандарт отрасли
- Очень неудобен
- Труден в освоении
- Стоимость/трудность взлома

5. R

- Может все
- Таблицы приличного вида сложно получить
- БЕСПЛАТЕН
- Развивается сообществом – большой потенциал
- Труден в освоении
- Удобен

6. STATA etc. (знаком мало)

**СПАСИБО ЗА
ВНИМАНИЕ!
ВОПРОСЫ?**

«Терминологические джунгли»

Table 23.1. *Terminology jungle*: after Gordis³.

Case-control study	=	Retrospective study
Cohort study	= Longitudinal study	= Prospective study
Concurrent cohort study	= Prospective cohort study	= Concurrent prospective study
Retrospective cohort study	= Historical cohort study	= Non-concurrent prospective study
Randomised trial	=	Experimental study
Cross-sectional study	=	Prevalence study

Introductory medical statistics, 3rd edition by R.F.Mould, 1998