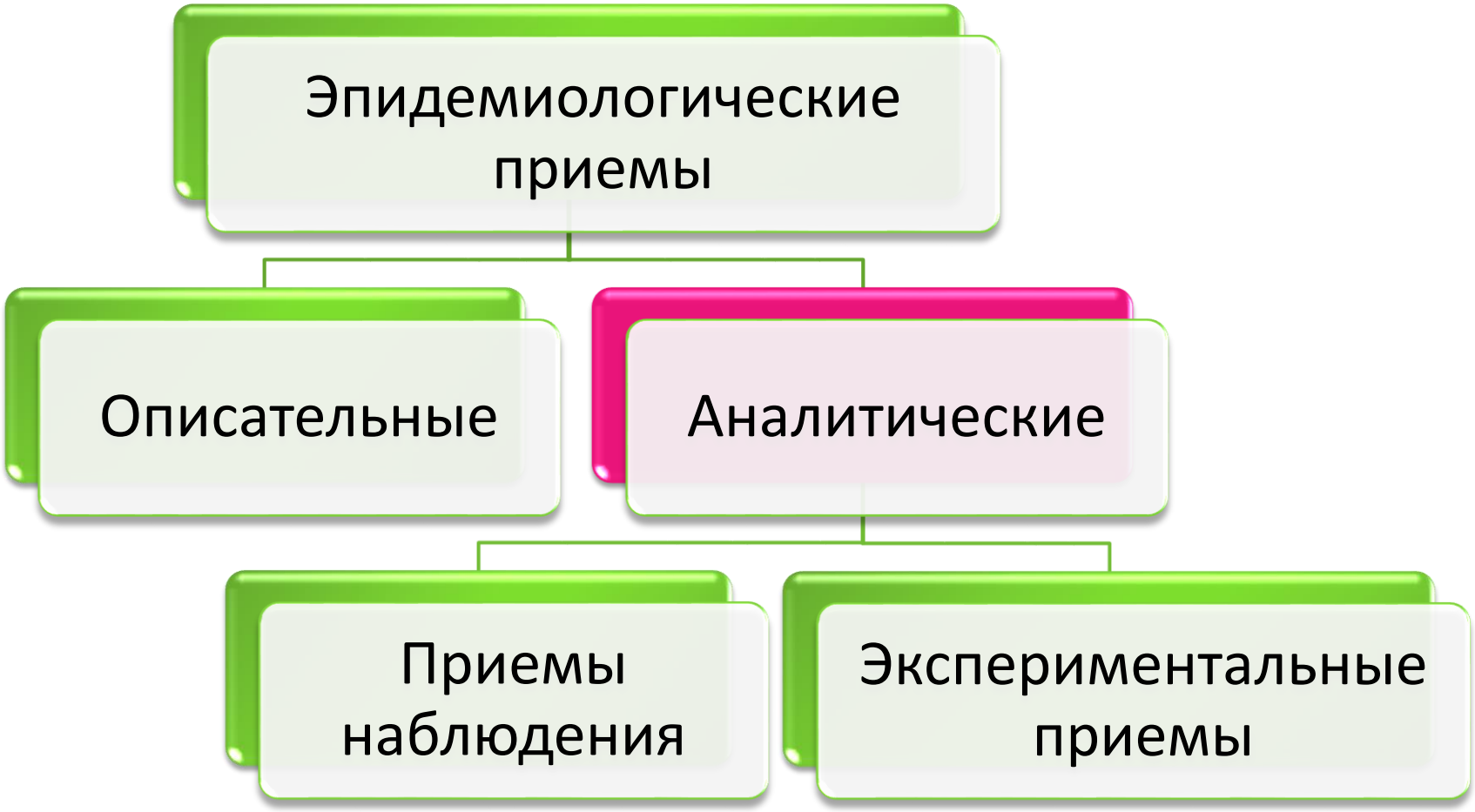
Эпидемиологические исследования. Потенциальные ошибки различных эпидемиологических исследований

Хасанова Г.Р.

Эпидемиологические приемы



Описательная эпидемиология

Описательные приемы

- Совокупность приемов, обеспечивающих сбор, обработку и интерпретацию данных о распространенности заболеваний и факторов риска в популяции
  - Во времени
  - В пространстве
  - В группах населения
- Пример: эпидемиологический надзор
- Помогают сформировать гипотезу исследования, отслеживать тенденции
- Являются основой для аналитической эпидемиологии

1. Эпидемиологическая диагностика / Л.П.Зуева и др. – 2-е изд. СПб: ООО «Издательство ФОЛИАНТ», 2009. (стр.12) – с дополнениями

Три ключевых вопроса

- Кто?
  - Группы людей по разным признакам или индивидуумы
- Где?
  - Территориальная распространенность
- Когда?
  - Распределение во времени и тенденции распределения

- Приемы описательной эпидемиологии дают представление об: 1) интенсивности, 2) динамике, 3) пространственной характеристике и
4) структуре изучаемого явления в популяции

Основные показатели описательной эпидемиологии

- Заболеваемость – показатель, характеризующий число новых случаев болезни (явления)
  - Инцидентность
- Распространенность – показатель, характеризующий общее количество существующих случаев
  - Превалентность

Показатели заболеваемости (инцидентность)

- Измерение частоты новых случаев (исходов) в популяции риска
  - Т.е. **только** среди тех, у кого вероятность развития исхода не равна нулю
- В аналитической эпидемиологии чаще всего применяются в контексте когортных исследований
- Основа для многих аналитических приемов:
  - Отношение риска
  - Разность риска
  - Анализ выживаемости и другие

- Эпидемиологическая диагностика / Л.П.Зуева и др. – 2-е изд. СПб: ООО «Издательство ФОЛИАНТ», 2009. (стр.58) – с дополнениями

Кумулятивная инцидентность (КИ)

- Общее число новых случаев заболевания (или иного исхода) в популяции риска
- Как правило, измеряется за определенный период времени (чаще – за год)
- Формула расчета:

$$КИ = \frac{\text{число новых случаев}}{\text{численность популяции риска}} \cdot 10^x$$

- Для удобства восприятия умножается на число, кратное 10 (100, 1000, 10.000, 100.000), так, чтоб результат был больше 1
  - Например: 5,5 на 1000 человек, а не 0,055 (на 1 человека)

Пример расчета кумулятивной инцидентности (1)

- Пример: орхит при эпидемическом паротите
- Проведите расчет показателя инцидентности для Республики Татарстан, если за год было зарегистрировано 3 случая орхита
- Общее население республики для примера: 4 млн. человек

Пример расчета кумулятивной инцидентности (1)

- 1 вариант:

$$КИ = \frac{3\ случая}{4.000.000} = 7{,}5\ на\ 10\ млн\ человек$$

- 2 вариант:

$$КИ = \frac{3\ случая}{4.000.000 \cdot 0{,}5} = 15\ на\ 10\ млн\ человек$$

- Какой вариант расчета вы выберете и почему?

Пример №2

- Рассчитайте, пожалуйста, инцидентность катетер-ассоциированного сепсиса в инфекционной больнице (за год)
- За год в больнице было установлено 500 катетеров и возникло 5 случаев катетер-ассоциированного сепсиса

Плотность инцидентности (риск)

- Кумулятивная инциндентность не учитывает продолжительность экспозиции (не учитывает длительность воздействия)
  - Человек, с продолжительностью катетеризации в 1 день уравнивается с тем, у кого катетер был установлен в течение 10 дней
- Плотность инцидентности учитывает продолжительность экспозиции (воздействия)

$$ПИ = \frac{\text{Число новых случаев}}{\text{Сумма времени риска}} \cdot 10^x$$

Пример расчета плотности инцидентности

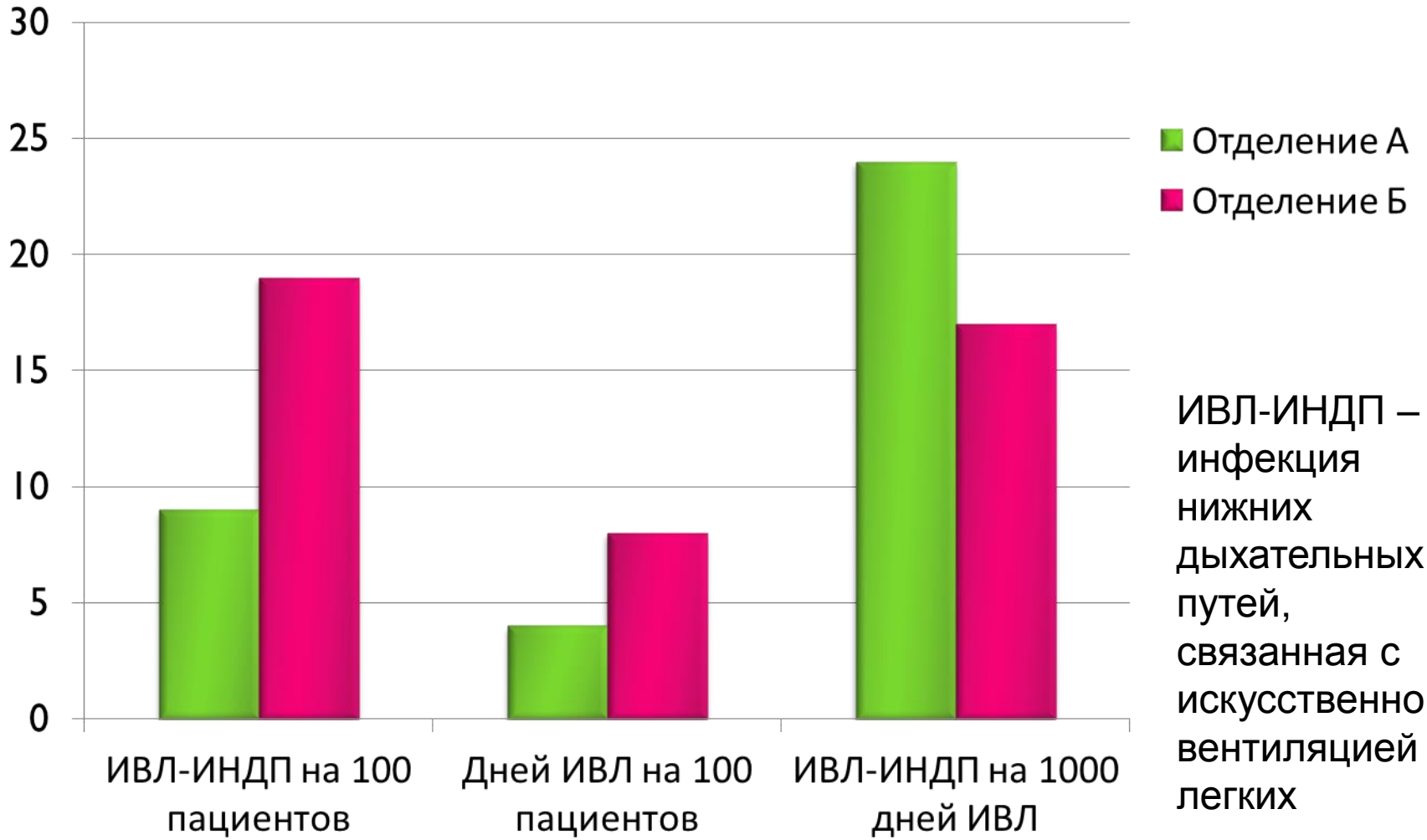
- Плотность инцидентности катетер-ассоциированного сепсиса:

$$ПИ = \frac{5\ \text{случаев}}{1\ \text{день} + 5\ \text{дней} + 3\ \text{дня} + 10\ \text{дней} + \ldots} =$$

$$= \frac{5\ \text{случаев}}{2000\ \text{дней катетеризации}} =$$

= 2,5 случая на 1000 дней катетеризации

Пример использования ПИ



ИВЛ-ИНДП – инфекция нижних дыхательных путей, связанная с искусственной вентиляцией легких

- Эпидемиологическая диагностика / Л.П.Зуева и др. – 2-е изд. СПб: ООО «Издательство ФОЛИАНТ», 2009. (стр.62)

Плотность инцидентности (2)

- Плотность инцидентности может использоваться для расчета риска развития исхода
- Плотность инцидентности предпочтительнее кумулятивной инцидентности, если продолжительность экспозиции имеет значение
  - В случае коротких экспозиций плотность инцидентности и кумулятивная инцидентность равны
- Плотность инцидентности позволяет оценивать заболеваемость в популяциях, размер которых меняется со временем
  - Для кумулятивной инцидентности нужна фиксированная популяция

Превалентность

- Количество случаев заболевания на определенный момент времени (моментная превалентность)
- Рассчитывается по формуле:

$$\text{Превалентность} = \frac{\text{Число имеющихся случаев}}{\text{Численность популяции риска}} \cdot 10^{x}$$

- Превалентность чаще всего используется в рамках срезовых исследований

Описательная статистика

Виды эпидемиологических данных

- Пол
- Национальность
- Группа крови

- Стадия заболевания
- Тяжесть состояния

- Количество детей
- Количество беременностей
- Возрастная группа

- Рост
- Вес
- Возраст
- Уровень холестерина крови

Особая группа – бинарные данные (дихотомические переменные)

Эпидемиологическая диагностика / Л.П.Зуева и др. – 2-е изд. СПб: ООО «Издательство ФОЛИАНТ», 2009. (стр.52) – с дополнениями

Количественные данные в эпидемиологии

- Для описания популяции зачастую необходимо использование количественных данных:
  - Артериальное давление
  - Уровень гемоглобина
  - Рост
  - Вес
  - Возраст
- Для этого используются:
  - Меры центральной тенденции
  - Меры рассеяния

Меры центральной тенденции

- Меры центральной тенденции характеризуют наиболее типичное для выборки значение
  - Среднее арифметическое (M)
    - Иначе – математическое ожидание
  - Медиана (Me) – показатель делящий ряд индивидуальных значений на две равные части (середина ранжированного ряда)
  - Мода – наиболее часто встречающееся значение признака в выборке

Среднее арифметическое

- Формула:

$$M = \frac{\text{сумма всех наблюдений}}{\text{число наблюдений}}$$

- Преимущества:
  - Существует для любых вариационных рядов
  - Всегда уникальна
  - Представляет широкие возможности для дальнейшего статистического анализа
  - Можно объединять несколько средних в одно общее значение
  - Легко рассчитать
  - Сравнительно надежна
- Недостаток: учитывает все значения

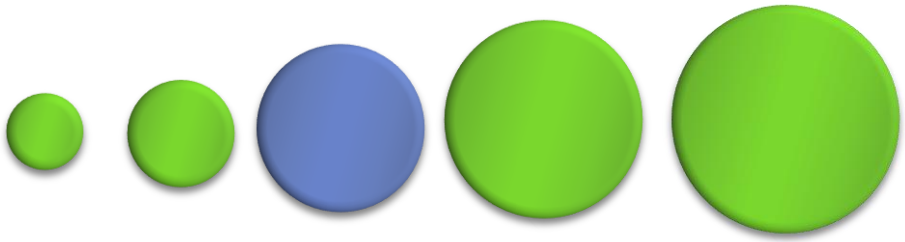
Пример расчета средней арифметической

	Должность	Зарплата (р/мес)
1	Директор	100.000
2	Помощник директора	50.000
3	Младший помощник директора	40.000
4	Старший работник 1	10.000
5	Старший работник 2	10.000
6	Работник	8.000
7	Младший работник 1	3.000
8	Младший работник 2	3.000
9	Младший работник 3	3.000
10	Младший работник 4	3.000

Средняя зарплата в трудовом коллективе = 22.000 рублей в месяц

Медиана

- Среднее значение ранжированного ряда
  - Для того, чтоб рассчитать медиану, нужно выстроить данные в порядке возрастания (или убывания)
  - Медиана – число, оказавшееся в середине ряда



Найдите медиану в ряду:
1 ряд: 1, 3, 4, 7, 10, 25, 32
2 ряд: 3, 7, 9, 11, 23, 43, 55, 67

Свойства медианы

- Преимущества:
  - Всегда существует
  - Всегда уникальна
  - Возможно дальнейшее использование в некоторых статистических приемах
  - Мало зависит от крайних значений
  - Легко рассчитать, если число наблюдений не велико
- Недостатки
  - Если число наблюдений велико – рассчитать медиану не просто
  - Нельзя объединять несколько медиан в одно значение
  - Не столь надежна, как среднее арифметическое

Пример расчета медианы

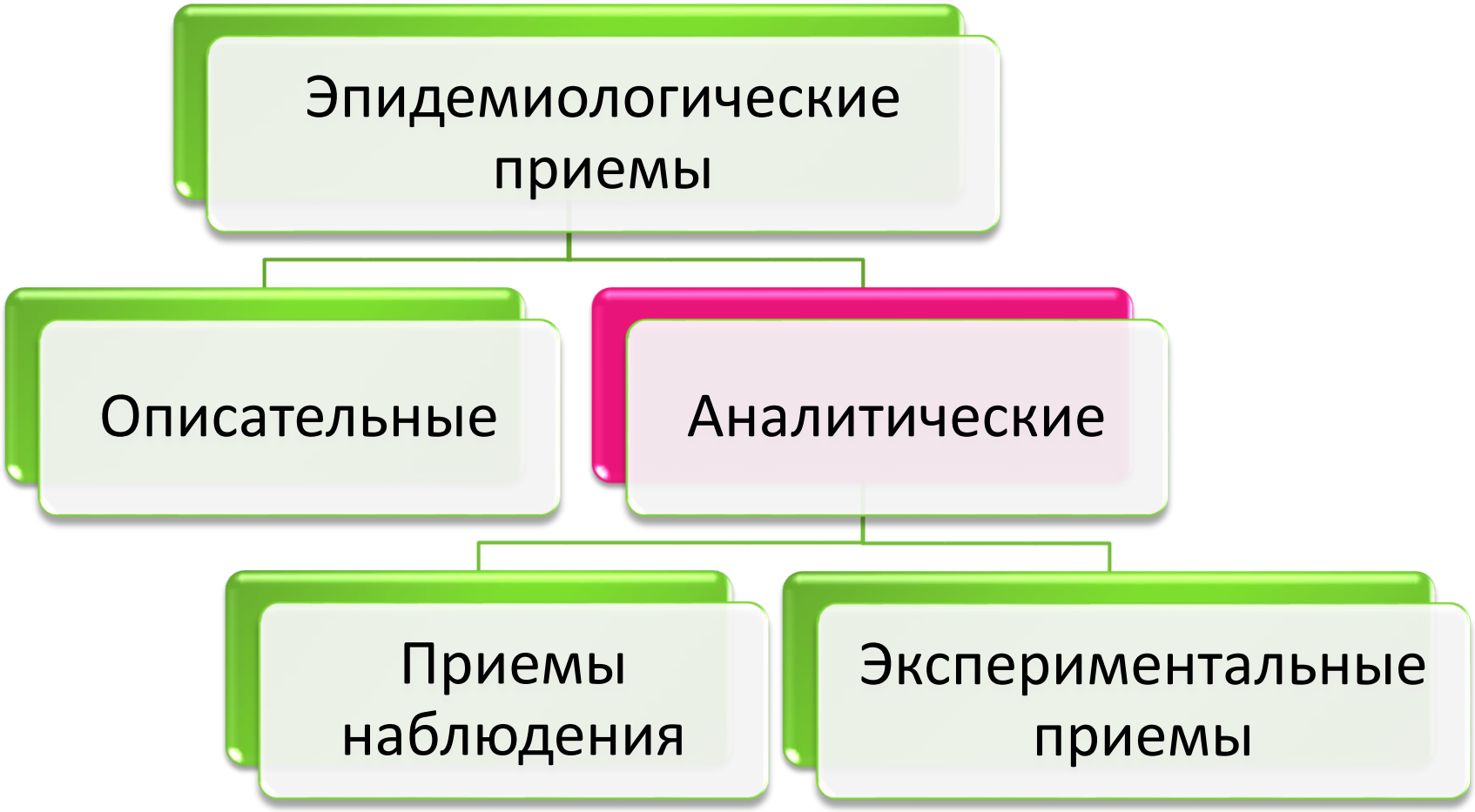
	Должность	Зарплата (р/мес)
1	Директор	100.000
2	Помощник директора	50.000
3	Младший помощник директора	40.000
4	Старший работник 1	10.000
5	Старший работник 2	10.000
6	Работник	8.000
7	Младший работник 1	3.000
8	Младший работник 2	3.000
9	Младший работник 3	3.000
10	Младший работник 4	3.000

Медианная зарплата в трудовом коллективе = (8.000 + 10.000)/2 = 9.000 рублей

Мода

- Наиболее часто встречающееся значение вариационного ряда
- Преимущества: может использоваться для работы с качественными данными
  - В том числе – для номинальных данных
- Недостатки:
  - Не всегда существует
  - Не всегда уникальна
  - Статистическая обработка не возможна
  - Моды нельзя объединять

Эпидемиологические приемы

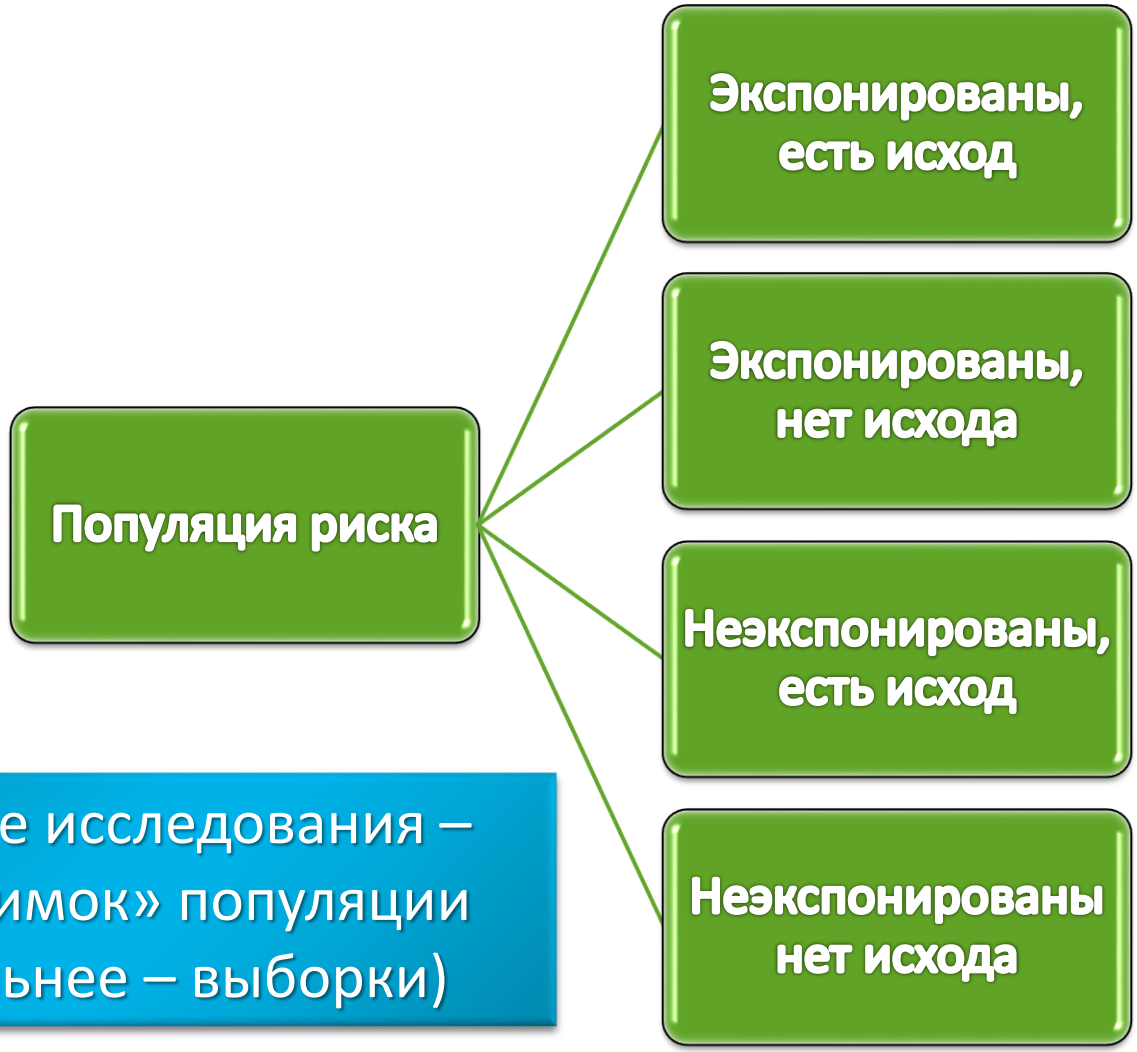


Аналитическая эпидемиология

- Аналитическая эпидемиология – это комплекс приемов, методов и подходов, направленных на оценку гипотез о причинах и условиях возникновения заболеваний (других исходов)

- Задачи аналитических исследований:
  - Измерение эффекта воздействия фактора
  - Оценка силы связи
  - Проверка причинности выявленных ассоциаций

Срезовые исследования



Срезовые исследования – «фотоснимок» популяции (правильнее – выборки)

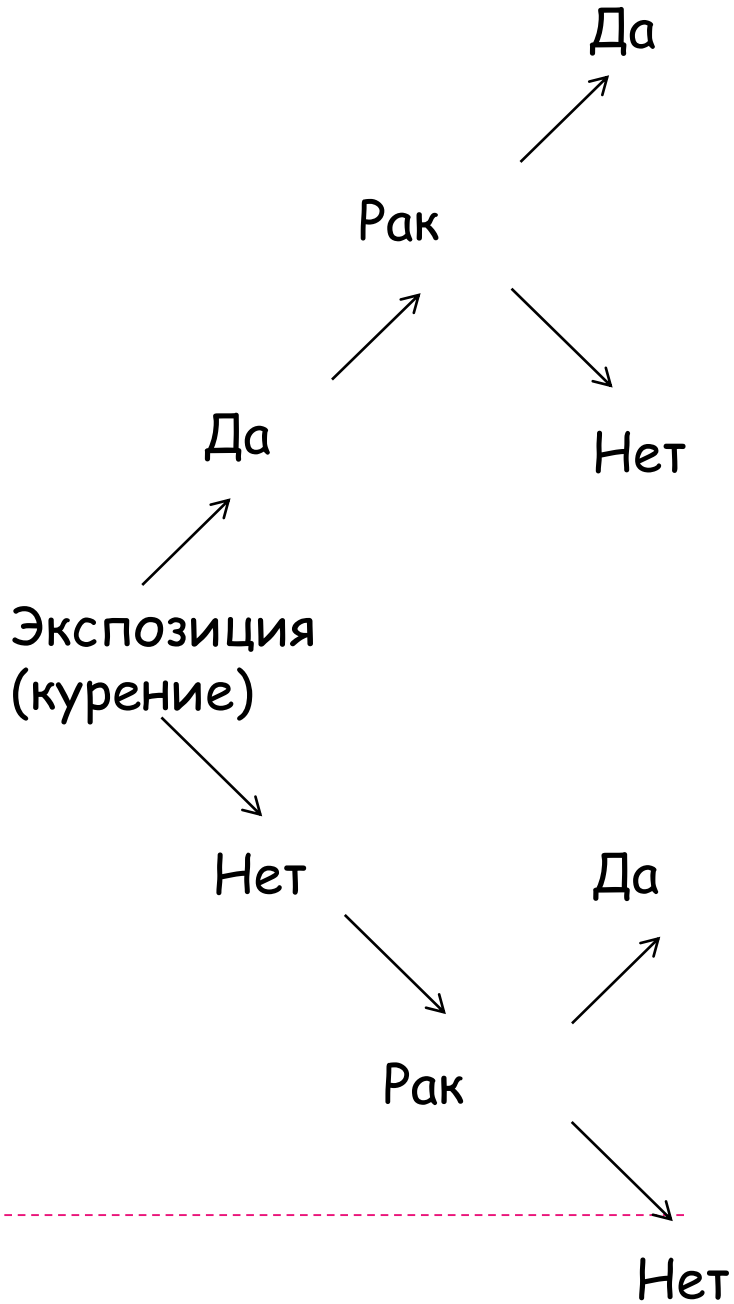
Основные преимущества и особенности срезовых исследований

- Как правило, самые недорогие и самые быстрые
- Нет проблемы потери участников
- Распространенность фактора риска оценивается одновременно с распространенностью исхода
  - Не всегда имеет смысл
  - Невозможна оценка временных взаимосвязей
  - Могут проводится серии срезовых исследований

Когортные исследования

Проспективное когортное исследование

- Изначально – 2 когорты – курящие и некурящие
- Наблюдение за группами для того, чтобы выяснить, у кого из включенных в когорту разовьется рак, у кого – нет.



Рак и курение, часть 2
Doll и Hill, 1954

- К участию приглашены 59,600 членов Британской медицинской ассоциации.
- Сбор данных:
  - В 1951 раздали короткий опросник с вопросами о табакокурении
  - Анкетируемые были поделены на группы после получения результатов
  - В 1954 (через 29 месяцев) из общей базы данных были взяты данные о смертности

Результаты

TABLE IV.—*Standardized Death Rate Per Annum Per 1,000 Men Aged 35 Years and Above in Relation to the Most Recent Amount of Tobacco Smoked*

Cause of Death	No. of Deaths Recorded	Death Rates of Non-smokers	Death Rates of Men Smoking a Daily Average of			Death Rate of All Men
			1 g.–	15 g.–	25 g.+	
Lung cancer	36*	0·00	0·48	0·67	1·14	0·66
Other cancers	92	2·32	1·41	1·50	1·91	1·65
Respiratory disease (other than cancer)	54	0·86	0·88	1·01	0·77	0·94
Coronary thrombosis ..	235	3·89	3·91	4·71	5·15	4·27
Other cardiovascular diseases	126*	2·23	2·07	1·58	2·78	2·14
Other diseases	247	4·27	4·67	3·91	4·52	4·36
All causes	789	13·61	13·42	13·38	16·30	14·00

* 1 case in which lung cancer was recorded as a contributory but not a direct cause of death has been entered in both groups.

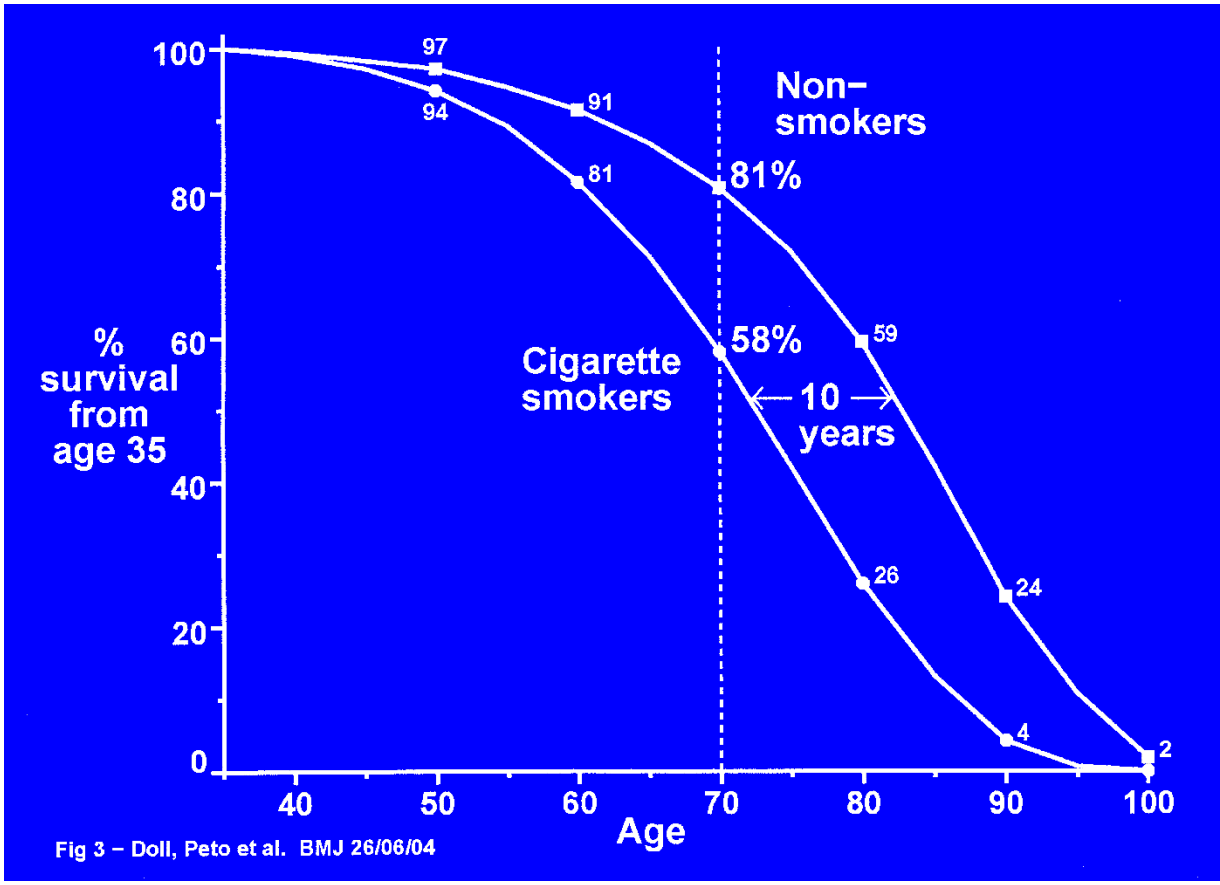
- Более высокие показатели смертности от рака легкого и ИБС среди курящих (а также тромбоза коронарных артерий)
- Более высокие показатели общей смертности среди курящих.

Результаты

- Преимущества когортного исследования
- Ограничения:
  - Малое количество смертей
  - Короткий период наблюдения
  - Doll и Hill сочли результаты предварительными

Продолжительное проспективное когортное исследование, Doll и Peto, 2004

- Доктора-мужчины (Великобритания)
  - 1900-1930 г.р.
  - 50 лет наблюдения 1951-2001



Четырехпольная таблица

- Данные многих эпидемиологических исследований сводятся в четырехпольной таблице (таблица 2х2)
  - Вариант 2 х N – если имеется несколько уровней фактора риска или несколько градаций исхода

	Исход имеется	Исхода нет	
Экспонированы	a	b	a + b
Неэкспонированы	c	d	c + d
	a + c	b + d	N = a + b + c + d

- Эпидемиологическая диагностика / Л.П.Зуева и др. – 2-е изд. СПб: ООО «Издательство ФОЛИАНТ», 2009. (стр.103)

Абсолютный риск (инцидентность)

	Исход имеется	Исхода нет	
Экспонированы	a	b	a + b
Неэкспонированы	c	d	c + d
	a + c	b + d	N = a + b + c + d

- R_e – вероятность возникновения изучаемого исхода у лиц, подверженных воздействию определенного фактора (у экспонированных)
- R_{ne} – риск исхода среди неэкспонированных
- R_e= a/a+b
- R_{ne} = c/c+d
- Диапазон [0;1]
- Абсолютный риск = инцидентность?

Относительный риск (отношение рисков, отношение инцидентностей)

	Исход имеется	Исхода нет	
Экспонированы	a	b	a + b
Неэкспонированы	c	d	c + d
	a + c	b + d	N = a + b + c + d

- Относительный риск (RR) – отношение абсолютных рисков, т.е. он показывает, во сколько раз риск заболевания в группе, подверженной воздействию, выше в сравнении с группой, не подверженной воздействию.
- Не используется в исследованиях «случай-контроль»!

$$RR = \frac{Re}{Rne} = \frac{a / (a + b)}{c / (c + d)}$$

RR

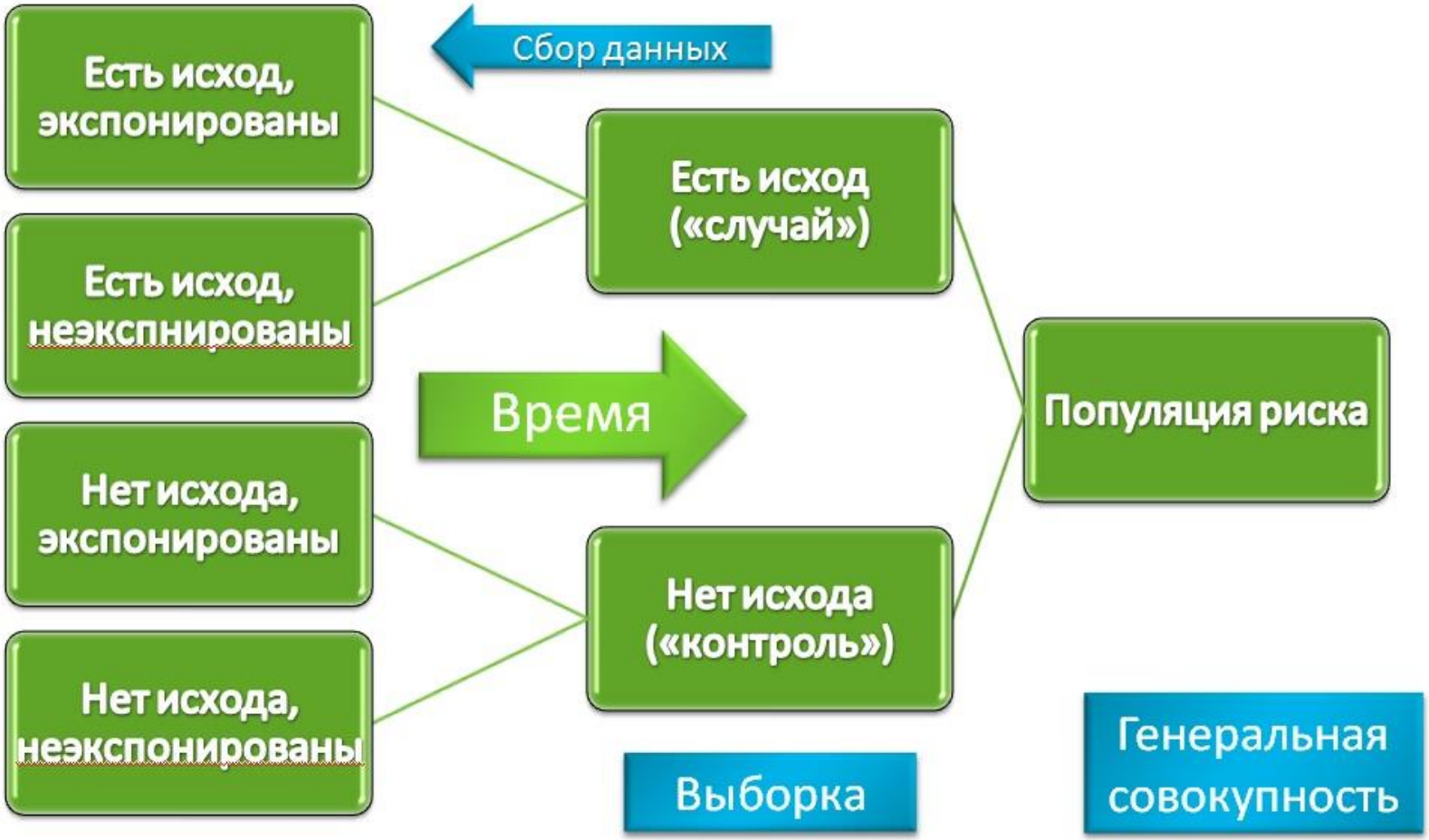
RR > 1.0 – экспозиция ассоциируется с повышением риска

RR = 1.0 - экспозиция не влияет на риск

RR < 1.0 – экспозиция асоциируется со снижением риска (протективный эффект воздействия)

Диапазон?

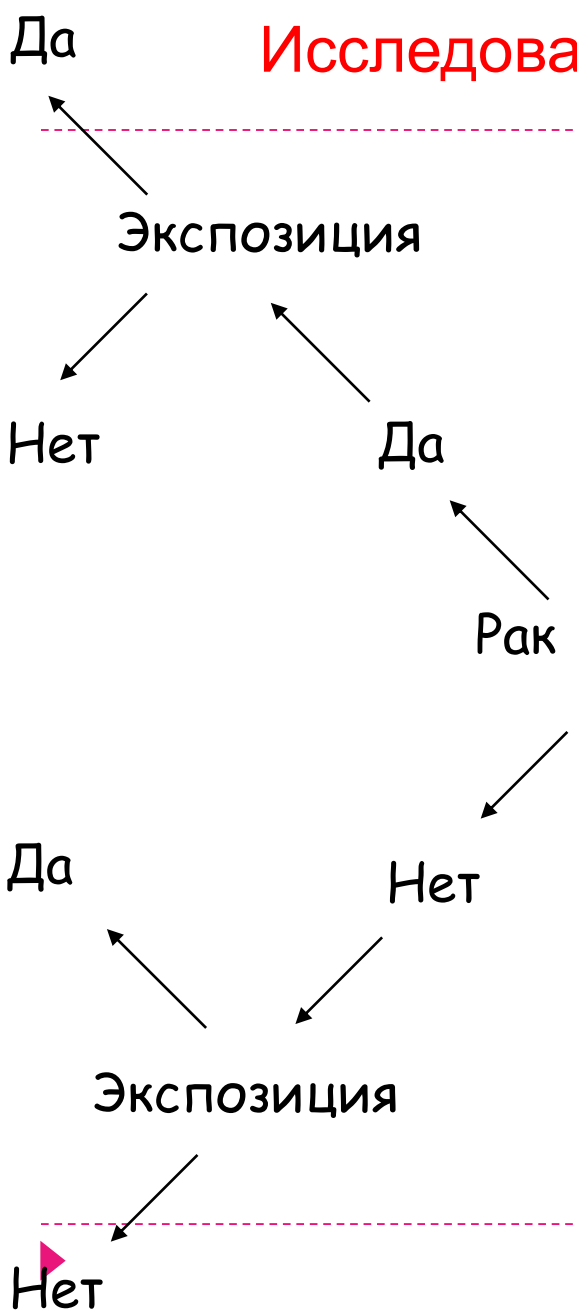
Случай-контроль



Курение и рак легкого: Doll и Hill, 1950

- *“The present investigation was planned in 1947 to be carried out on a sufficiently large scale to determine whether patients with carcinoma of the lung differed materially from other persons in respect of their smoking habits or in some other way which might be related to the atmospheric pollution theory.”*
- «Настоящее исследование было спланировано в 1947 г. для того, чтобы выяснить, отличаются ли пациенты с раком легкого от других людей по курению табака или экспозиции к другим атмосферным загрязнениям

Исследование «случай-контроль»: взгляд назад



- Изначально берется группа людей с раком.
- Подбирается контрольная группа.
- В обеих группах собирается анамнез, касающийся курения

Исследование случай-контроль: методика

- Набраны группы «случая» и «контроля» -- в зависимости от факта наличия или отсутствия заболевания (не фактора риска!)
  - Группа «случая»: 709 больных раком легкого
  - Группа «контроля»: 709 больных с другими диагнозами
  - Группы сопоставимы по полу, возрасту и госпиталю
- Сбор данных: интервью
  - Курил ли когда либо, возраст начала/прекращения, количество сигарет, трубка или сигареты, затягивается или нет
  - Курящими считались люди, выкуривающие хотя бы одну сигарету в день в течение 1 и более лет
- Анализ: сравнение доли курящих (с учетом количества сигарет, продолжительности курения, других привычок) в группах «случая» и «контроля».

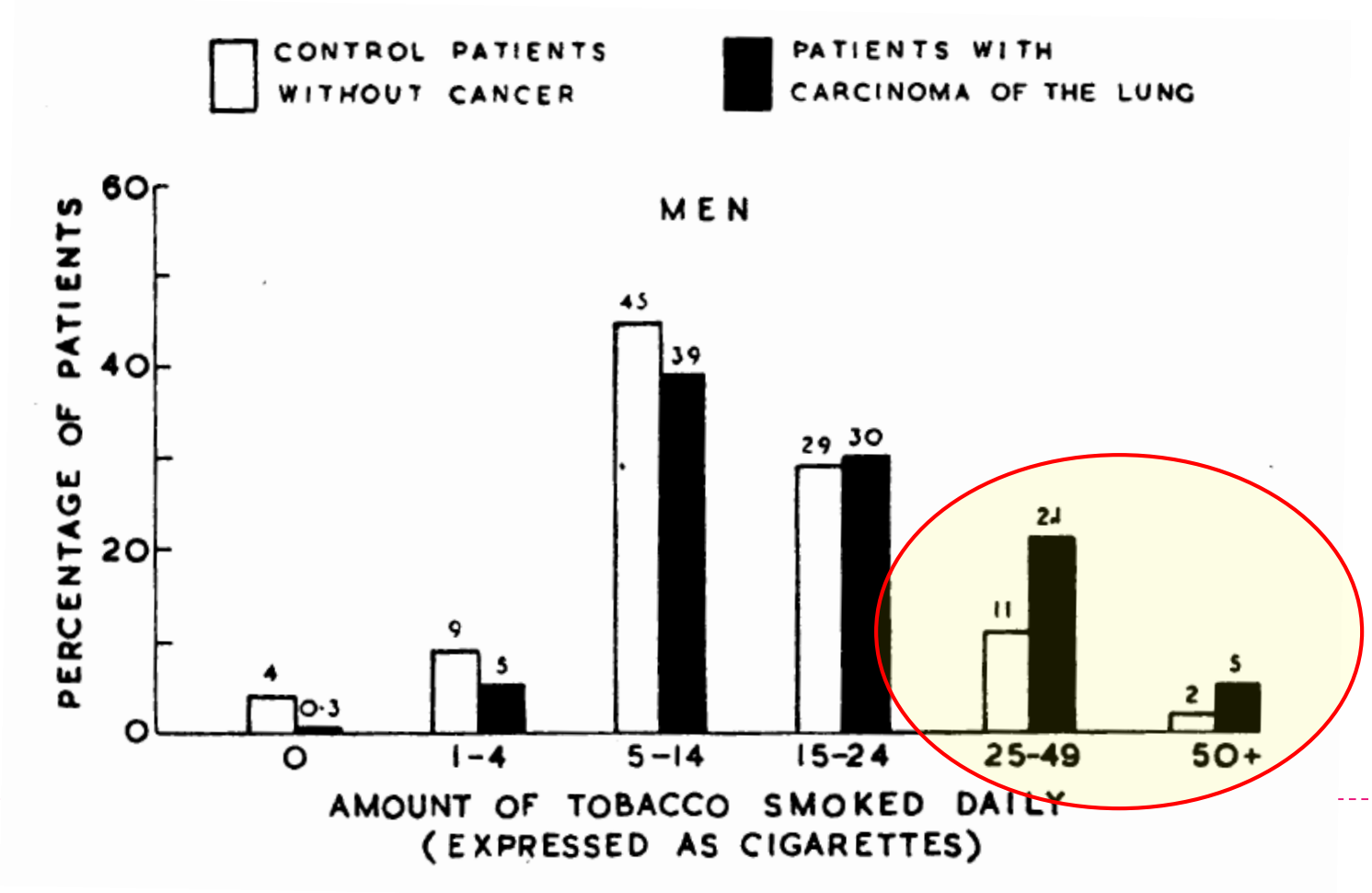
Результаты:
“”

TABLE IV.—*Proportion of Smokers and Non-smokers in Lung-carcinoma Patients and in Control Patients with Diseases Other Than Cancer*

Disease Group	No. of Non-smokers	No. of Smokers	Probability Test
Males:			
Lung-carcinoma patients (649)	2 (0·3%)	647	P (exact method) = 0·00000064
Control patients with diseases other than cancer (649) ..	27 (4·2%)	622	
Females:			
Lung-carcinoma patients (60)	19 (31·7%)	41	$\chi^2 = 5{\cdot}76$; n = 1 $0{\cdot}01 < P < 0{\cdot}02$
Control patients with diseases other than cancer (60) ..	32 (53·3%)	28	

- Большая часть пациентов курили (только 70 из 709 - некурящие)
- Среди женщин доля курящих меньше (42,5% против 97,8% у мужчин)
- Значительно меньшее количество рака легких у женщин (60) в сравнении с мужчинами (649).

Зависимость от количества сигарет



Заключение

- “Курение – важный фактор риска в развитии рака легкого”

Отношение шансов (odds ratio,OR)

- Исследования «случай-контроль» – только отношение шансов (отношение вероятностей)!
- Шанс = отношение количества экспонированных к неэкспонированным среди случаев или контролей
- **OR = a/c : b/d = ad/bc**

	Исход имеется (случай)	Исхода нет (контроль)	
Экспонированы	a	B	a + b
Неэкспонированы	c	d	c + d
	a + c	b + d	N = a + b + c + d

Необходимо помнить об особенностях интерпретации результатов исследований «случай-контроль»!

Эпидемиологическая диагностика / Л.П.Зуева и др. – 2-е изд. СПб: ООО «Издательство ФОЛИАНТ», 2009. (стр.103-105) – с дополнениями

Отношение шансов (odds ratio, OR)

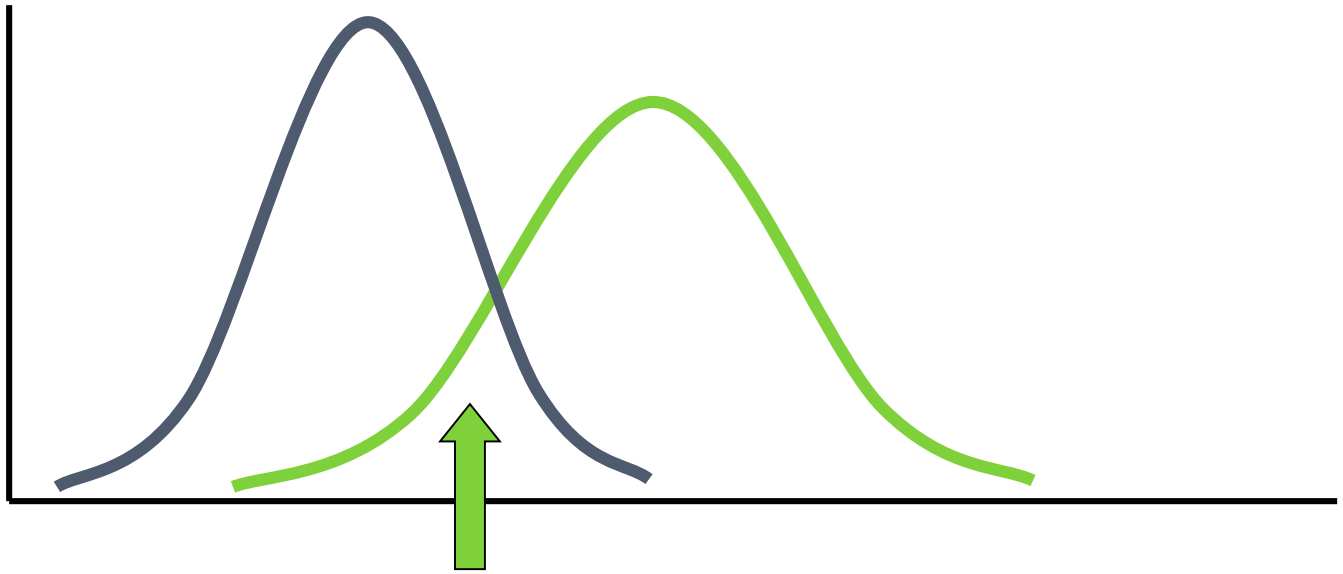
OR > 1.0 экспозиция ассоциируется с повышенным риском исхода

OR = 1.0 экспозиция не влияет на риск исхода

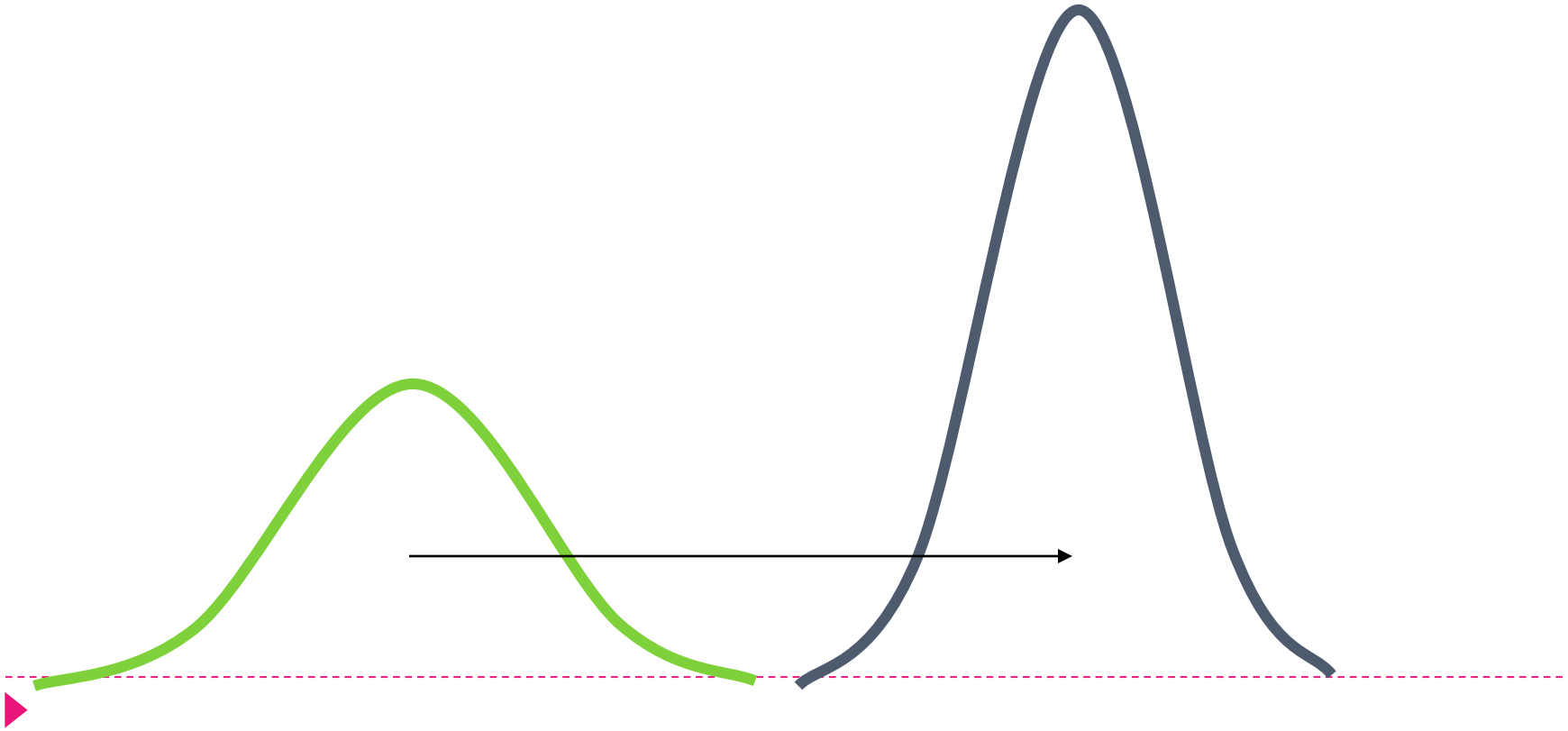
OR < 1.0 экспозиция ассоциируется со снижением риска развития исхода (протективный эффект)

Возможный диапазон?

Гипотетическое клиническое исследование:



«нахлест» <5%?



- Проверка нулевой гипотезы.
- $\lambda^2 + p$ или критерий Фишера?
- Доверительный интервал

Хи-квадрат-тест

Метод статистического анализа для сравнения двух и более пропорций при большом числе наблюдений (в данном случае, сравнение исходов среди экспонированных и неэкспонированных).

$\lambda^2 = (ad-bc)^2 \times N / (a+c)(b+d)(a+b)(c+d)$

При $\lambda^2 > 3{,}84$ $p < 0{,}05$

Метод Фишера

Статистический анализ для оценки различий между 2 и более пропорциями при малом числе наблюдений (<5 хотя бы в одной клетке)

	Исход имеется	Исхода нет	
Экспонированы	30	60	90
Неэкспонированы	1	29	30
	21	79	N = 120

- Бесплатный пакет EpiInfo можно загрузить с адреса: http://wwwn.cdc.gov/epiinfo/html/downloads.htm
- или http://www.openepi.com

Корреляционный анализ

- Оценка взаимосвязи количественных или качественных порядковых данных
- Коэффициент корреляции показывает, в какой мере изменение значения одной переменной сопровождается изменением значения другой переменной в конкретной популяции
- Мера – коэффициент корреляции ***r***
  - Диапазон значений от -1 до +1
  - 0 означает отсутствие взаимосвязи
  - Положительные значения – прямая взаимосвязь, отрицательные – обратная

Корреляционный анализ (2)

- Не может служить для оценки причинно-следственных взаимосвязей
  - Корреляция может означать, что:
    - Признак А влияет на признак Б
    - Признак Б влияет на признак А
    - Оба признака одновременно находятся под влиянием третьих факторов
- Условно, принято считать, что:
  - $r < 0{,}25$ – слабая корреляция
  - $0{,}25 < r < 0{,}75$ – корреляция средней силы
  - $r > 0{,}75$ – сильная корреляция

Тестирование гипотезы в корреляционном анализе

- Нулевая гипотеза – изучаемые факторы не взаимосвязаны
- Значение ***p*** отражает вероятность альфа-ошибки
  - Вероятность, что наблюдаемое значение является результатом статистической аберрации
- Пример: выявлена корреляционная взаимосвязь между температурой тела и уровнем эндотоксина в крови (r=0,57; p=0,02)

Общие принципы организации проведения клинических испытаний

- Контролируемость
- Рандомизированность
- Обязательное соблюдение всех принципов и этических норм, представленных в Хельсинкской декларации

Любые наблюдения подвержены влиянию **случайности.**

Случайная ошибка **(СлО)** – отклонение результата (отдельного) наблюдения в выборке от истинного значения в популяции, обусловленное исключительно случайностью.

Пример 1: подброшенная 100 раз монета падает орлом не точно 50 раз.

Пример 2: если предположить, что два препарата одинаково эффективны и каждый из них вызывает улучшение примерно у половины больных, то в единичном исследовании с небольшим числом больных в сравниваемых группах вполне может оказаться (исключительно из-за чистой случайности), что прием препарата А в большем проценте случаев дает положительный эффект, чем препарат Б (*Р.Флетчер, 1998)*.

СисО в клиническом наблюдении (основные категории):

***СисО**, обусловленная отбором* *(смещение выборки)*, возникает, когда сравниваемые группы пациентов различаются не только по изучаемому признаку, но и по другим факторам, влияющим на исход.

Пример 1: На этапе отбора в исследование, пациентам с исходно более тяжелыми проявлениями болезни назначается препарат А (новое, очень популярное средство), а не Б (старое средство).

СисО, *обусловленная отбором (смещение выборки)*

Пример 2: в когортных исследованиях по изучению рисков и пользы от употребления алкоголя, была выявлена «J-образная» связь между употреблением алкоголя и смертностью. Лучший исход (по показателям преждевременной смерти) отмечался в группе умеренно пьющих людей. Трезвенники со значительно большей вероятностью погибали в более молодом возрасте, чем обычный человек, потреблявший умеренное количество алкоголя.

Почему был получен такой результат?

Популяция трезвенников включает:

- *людей, ведущих здоровый образ жизни;*
- *людей, вынужденных бросить пить по состоянию здоровья;*
- *тех, кто пьют «как рыбы», но предпочитают об этом не распространяться.*

(Т.Гринхальх «Основы доказательной медицины», 2006)

СисО в клиническом наблюдении (основные категории):

***СисО**, обусловленная измерением*, возникает, когда в сравниваемых группах больных используются разные методы измерения.

Пример 1: Изменения уровня АД в зависимости от условий измерения (*Sackett et al.*):

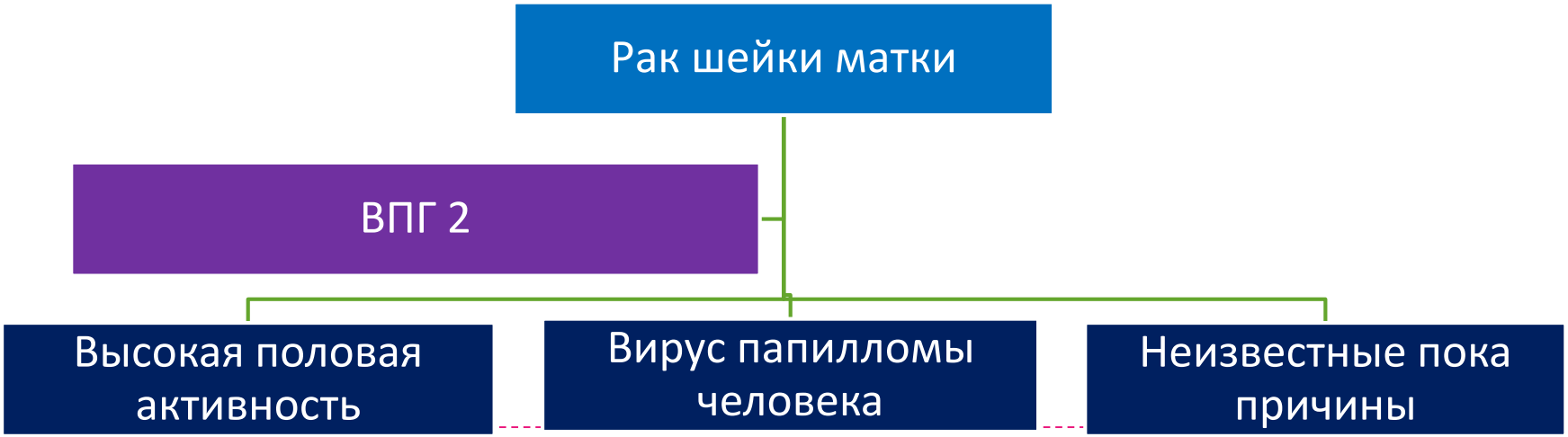
Измерение в кабинете, а не дома	+ 5 mm Hg
Незнакомый врач	+15
Разговоры	+10
Рука на весу	+8
Напряженная спина	+8
Послеобеденное время	+8

Пример 2: Частота применения контрацептивов у женщин, госпитализированных по поводу флеботромбоза и по другим причинам.

СисО в клиническом наблюдении (основные категории):

***СисО**, обусловленная вмешивающимися факторами (**конфаундинг**)* возникает, когда один фактор связан с другим, и эффект одного искажает эффект другого.

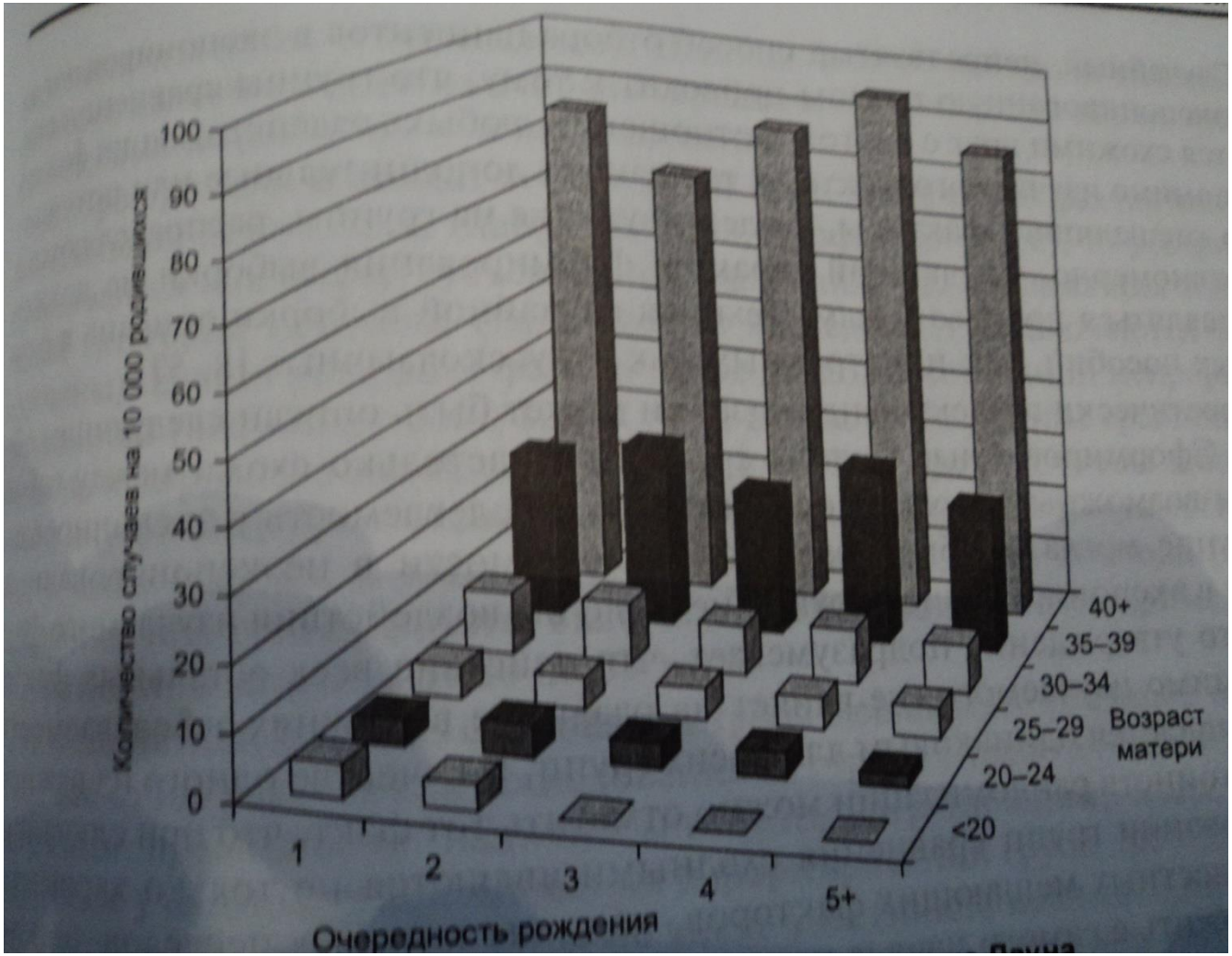
Пример 1:



Stark CR, Mantel N. Effects of maternal age and birth order on the risk of mongolism and leukemia. J Natl Cancer Inst., 1966; 37: 687-698.

Рис. 34. **Частота возникновения синдрома Дауна в зависимости от очередности рождения ребенка**

Количество случаев на 10 000 родов
90
80
70
60
50
40
30
20
10
0
<20
20–24
25–29
30–34
35–39
40+
Возраст матери


Количество случаев на 10 000 родившихся
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5+
Очередность рождения
40+
35–39
30–34
25–29
20–24
<20
Возраст матери

Способы контроля мешающих факторов

1. Рандомизация (равномерное распределение потенциальных мешающих факторов в сравниваемых группах путем их случайного формирования)
2. Рестрикция (исключение лиц с потенциальными мешающими факторами)
3. Подбор контролей
4. Стратификация (выделение страт, однородных с точки зрения мешающих факторов) с подсчетом взвешенного риска