

Некоторые проблемы анализа данных в медико- биологических исследованиях

О. Реброва

канд. мед. наук

НИИ неврологии РАМН

olga@neurology.med.ru

www.neurology.ru

(095) 490 2038

- **Биология – статистика не нужна:**

если Вам нужна статистика для анализа данных эксперимента, то Ваш эксперимент плох (слабый эффект, гетерогенность выборки).

- **Медицина – статистика нужна:**

поиск относительно небольших различий (обычно не более 20%) при значительных вариабельности данных и неточности измерений.

Доказательная медицина - клиническая эпидемиология

- Минимизация систематических ошибок вследствие отбора, измерения, вмешивающихся факторов
(адекватный дизайн исследования)
- Минимизация случайных ошибок (корректный статистический анализ)

Российское отделение Кокрановского
Сотрудничества - www.cochrane.ru

Что может статистика?

- **Статистическое оценивание**
- **Проверка гипотез**
- **Статистическое моделирование**

- *Придать исследованию наукообразность*

Что статистика не может?

- Улучшить выборку
- Оценить неизвестные признаки
- Исправить ошибки в измерениях
- Дать интерпретацию результатов

Почему статистика трудна для врачей?

- **Специальная терминология**
- **Необходимость абстрактного мышления**
- **Сложность вычислений**
- **Переоценка возможностей статистики**

Проблемы возникают на разных этапах исследования:

- I. Постановка задачи
- II. Подготовка данных к анализу
- III. Проверка данных
- IV. Выбор методов статистического анализа
- V. Интерпретация результатов
- VI. Представление результатов

I. Постановка задачи

- **Garbage in, garbage out**
- **Никакая статистическая обработка данных не может устранить неизвестную систематическую ошибку**
- **Проверка гипотез (первичный анализ данных) или выдвижение гипотез (вторичный анализ - post hoc analysis - data dredging)**

II. Подготовка данных

- **Разбиение области значений на интервалы, округление и точность**
- **Предварительные расчеты**
- **Использование стандартных шкал для клинических признаков**
- **Пропущенные значения**
- **Выбор объекта наблюдений**
- **Контрольные группы**
- **Интервал нормы**

III. Проверка данных

- **Ошибки набора**
- **Артефакты**
- **Выпадающие значения**

IV. Выбор методов статистического анализа

- **Типы данных**
- **Вид распределения**
- **Одно- и двусторонние тесты**
- **Связанные и несвязанные выборки**
- **Проблема множественных сравнений
(алгоритмы, выбор уровня P)**
- **Хи-квадрат или ТКФ**
- **Корреляция или регрессия**

V. Интерпретация результатов

- Отсутствие достоверных результатов не является подтверждением нулевой гипотезы
- Корреляционная связь – не причинно-следственная
- Валидизация многомерных моделей
- Data dredging (post hoc analysis)
- Соотношение статистической и клинической значимости
- Очень большие и очень маленькие выборки
- Суррогатные исходы и конечные точки

VI. Представление результатов

- «Единые требования к статьям, представляемым в международные биомедицинские журналы» (Межд. журнал мед. практики, 1997, N 5, с. 53-64)
- Число наблюдений для каждого признака
- Описательная статистика -
 $M \pm SD$, Me (LQ;UQ), % (n/N)
- Точность результатов (оценки, P)
- ДИ (для основных результатов исследования) и P
- Указание на использованные стат. методы
- Указание на использованный стат. пакет