

11-13 ДЕКАБРЯ 2025 · МОСКВА

XXIX РОССИЙСКИЙ ОНКОЛОГИЧЕСКИЙ КОНГРЕСС 2025



РАНДОМИЗИРОВАННЫЕ НЕСРАВНИТЕЛЬНЫЕ ИССЛЕДОВАНИЯ – ОКСЮМОРОН ИЛИ РЕАЛЬНАЯ ПОЛЬЗА *МНЕНИЕ СТАТИСТИКА*

Мироненко Ольга Николаевна

биостатистик, преподаватель Института биоинформатики

rosoncoweb.ru

С ЧЕГО ВСЁ НАЧАЛОСЬ

В том числе идея этой сессии

datamethods

- Topics
- More
- CATEGORIES
- general
- data analysis
- research methods
- journal club
- education
- All categories

Randomized non-comparative trials: oxymoron ?

interpretation design rct

terrible

disastrous

odd

2 Sep 2024

stupid

weird

unethical

WTF

schizophrenic

rubbish

statistical slop

Randomized non-comparative trials are becoming increasingly common, particularly in oncology, and being used by prominent clinicians. Randomized non-comparative trials (RNCTs) consist of two or more treatment arms and then *not* compare them to each other. Each arm is compared to a specified control. Essentially, RNCTs act as single-arm trials. Convenience sampling is used, similarly to standard comparative randomized trials, so control of the sampling process to facilitate estimation of group-level survival outcomes in each treatment group).



RNST ПРОКРАЛИСЬ НЕЗАМЕТНО...

Статистики всё проспали ☹️

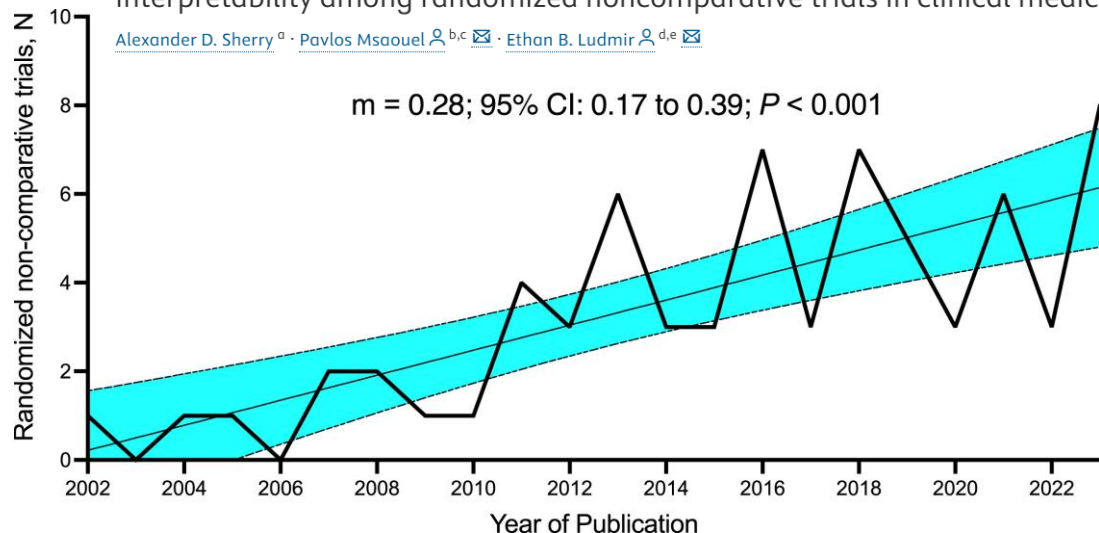
jce
Journal of
Clinical
Epidemiology

ORIGINAL RESEARCH · Volume 175, 111540, November 2024

Download Full Issue

A meta-epidemiological analysis of post-hoc comparisons and primary endpoint interpretability among randomized noncomparative trials in clinical medicine

Alexander D. Sherry^a · Pavlos Msaouel^{b,c} · Ethan B. Ludmir^{d,e}



RNST ПРОКРАЛИСЬ НЕЗАМЕТНО...

Статистики всё проспали 😞

SIGNIFICANCE



Pavlos_Msaouel

Apr 25

Since the article was published I have certainly received a fair number of messages and phone calls by statisticians in shock and unable to contain their use of 4-letter words 😊

Issues Advance Articles Submit ▼ Purchase



Volume 22, Issue 3
May 2025

JOURNAL ARTICLE

Bad stats: A regular series exploring slip-ups, snafus and salutary lessons from the world of statistics FREE

Pavlos Msaouel

Significance, Volume 22, Issue 3, May 2025, Pages 40–44, <https://doi.org/10.1093/jrssig/qmaf029>

Published: 28 March 2025



Кто виноват?*



Do we know the first researcher who used or proposed this method? I would love to know who to blame.



davidcnorrismd

Sep 2024

Is there some canonical set of methodological references on RNCT's, Pavlos? Or is this more an accidental development? Who are the statisticians supporting these designs?

С ЧЕГО ВСЁ НАЧАЛОСЬ НА САМОМ ДЕЛЕ?

Эффект «глухого телефона»



f2harrell

Sep 2024

There is a general theme that bothers me to no end. Statisticians often appear to prioritize helpfulness over principles.



Volume 22, Issue 3

May 2025

Misconceptions and a “broken telephone” effect

How did we get here? One plausible explanation is a kind of “broken telephone” in clinical research: a design is proposed, someone endorses it for reasons that might be poorly understood, and the next wave of researchers follows suit without stepping back to question the fundamentals. The impetus may be a widely held but vague notion that “randomisation is good”. If randomisation is viewed as a ritual that grants an “evidence-based medicine” halo, then we end up with RNCTs.

Citations to date

0

Altmetric

Elisabeth Coart & Everardo D. Saad

Pages 271-280 | Received 08 Feb 2021, Accepted 08 Apr 2021, Accepted author version posted online: 09 Apr 2021, Published online: 19 Apr 2021

Cite this article

<https://doi.org/10.1080/23808993.2021.1915693>



ПРИЧИНЫ РАСПРОСТРАНЕНИЯ RCT

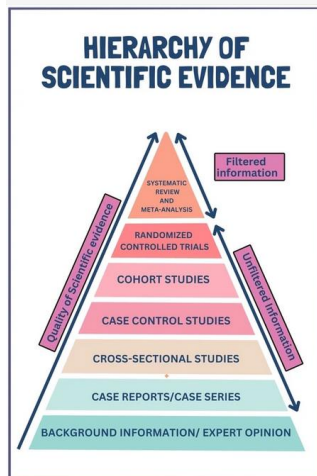
Гипотеза 1

- Рандомизация – «золотой стандарт»: + 100 к доказательности, + 100 к публикабельности

ПЛЮШКИ РАНДОМИЗАЦИИ

Необходимое и достаточное условие высокого уровня доказательности (НЕТ)

THE LANCET



[This journal](#) [Journals](#) [Publish](#) [Clinical](#) [Global health](#) [Multimedia](#) [Events](#) [About](#)

ARTICLES · Volume 403, Issue 10431, P1061-1070, March 16, 2024

[Download Full Issue](#)

Sunitinib for metastatic progressive pheochromocytomas and paragangliomas: results from FIRSTMAPPP, an academic, multicentre, international, randomised, placebo-controlled, double-blind, phase 2 trial

[Eric Baudin, MD](#) ^a [Prof Bernard Goichot, MD](#) ^b · [Prof Alfredo Berruti, MD](#) ^c · [Julien Hadoux, MD](#) ^a · [Salma Moalla, MD](#) ^a · [Sandrine Laboureau, MD](#) ^d · et al. [Show more](#)

[Affiliations & Notes](#) [Article Info](#) [Linked Articles \(1\)](#)

ПРИЧИНЫ РАСПРОСТРАНЕНИЯ RCT

Гипотеза 2

- Рандомизация – «волшебная палочка»: уничтожает все систематические ошибки, +100 к мощности

РАНДОМИЗИРОВАТЬ И НЕ СРАВНИВАТЬ

Но если очень хочется сравнить, то можно (НЕТ)

jce
Journal of
Clinical
Epidemiology

A meta-epidemiological analysis of post-hoc comparisons and primary endpoint interpretability among randomized noncomparative trials in clinical medicine

Alexander D. Sherry^a · Pavlos Msaouel^{b,c} · Ethan B. Ludmir^{d,e}

- After characterizing all published randomized noncomparative trials (RNCTs) in the medical literature, we find that 51% of RNCTs compared outcomes between arms, despite their study name, and 86% of RNCT primary outcomes lacked interpretability.

ПРИЧИНЫ РАСПРОСТРАНЕНИЯ RNCT

Гипотеза 3

- Большая вариабельность исходов / редкие события / небольшой предполагаемый эффект терапии:
 - +100500 к численности пациентов, которых нужно включить в РКИ
 - +100500 лет к длительности наблюдения за ними в РКИ
- В RNCT не предполагается сравнение групп рандомизации:
 - +100 к облегчению от отсутствия необходимости рассчитывать выборку для этой задачи
 - Или +100 к гибкости (замена на расчет выборки под сравнение среднего значения исхода с некоторым порогом)

МЕЖДУ СРАВНИТЕЛЬНЫМ И ОДНОРУКАВНЫМ

«Интересное решение» (HET)



simongates

Apr 17

The second of those trials (the Italian one) suggests a possible reason for using this design: they can compare each group to a single point value (and choose a nice low one) so more likely to get a “statistically significant” result.



simongates

Apr 18

Good question! There's definitely an ethical dimension here - if we're asking patients to join an experiment and accept a novel treatment that could harm them, we have an obligation to make sure that it's soundly designed and will yield useful information.

expected difference between arms, but simply on the expected results for the experimental arm, as in the single-arm design, with the advantage that the control arm will be used to assess the reliability of the historical data (i.e. of the null hypothesis) used in the design

А что, собственно, не так?



simongates

It's like people have forgotten what the whole point of randomisation is.



f2harrell

So we have
bad news → coverup → find a dishonest unethical design hiding the bad news



f2harrell

The authors' "calibration" remark is an example of ill-specified methods and fuzzy thinking.
Calibration in this context means anything you want it to mean.

ИССЛЕДОВАТЕЛЬСКИЙ ВОПРОС

Начало начал

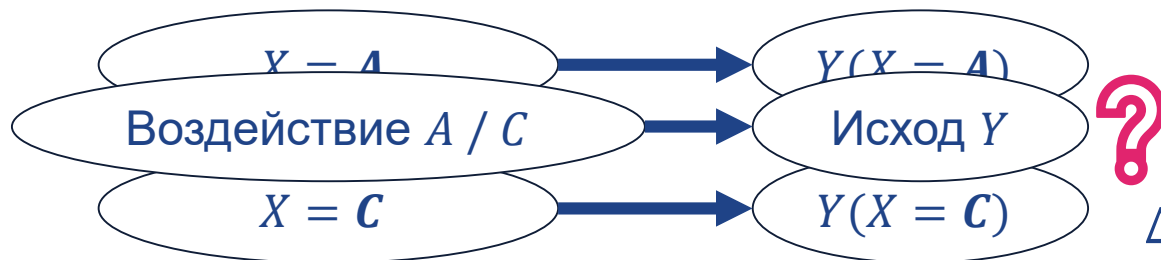
Исследовательский вопрос	Оценка	Проверка гипотезы
Описательный: охарактеризовать распределение показателя в целевой популяции	Частота объективного ответа (ЧОО) на терапию B у пациентов с заболеванием N с 95% ДИ	Исследовательская гипотеза: у пациентов с заболеванием N ЧОО на терапию $B > 40\%$ (сравнение с порогом)
Каузальный (причинно- следственный): оценить эффект (влияние) воздействия на исход в целевой популяции	Отношение рисков (HR) для общей выживаемости (ОВ) на терапии A по сравнению с терапией C у пациентов с заболеванием M	Исследовательская гипотеза: у пациентов с заболеванием M ОВ на терапии $A > ОВ$ на терапии C (сравнение с контрольной группой)
Предсказательный: предсказание исхода для новых наблюдений по их исходным характеристикам		

КАУЗАЛЬНЫЕ ИССЛЕДОВАНИЯ

Что было бы, если бы...?



- Оценить эффективность воздействия без контрольной группы невозможно
 - Что было бы с исходом, если бы пациент не получил это воздействие?



Эффект воздействия = $\overline{\Delta Y}$

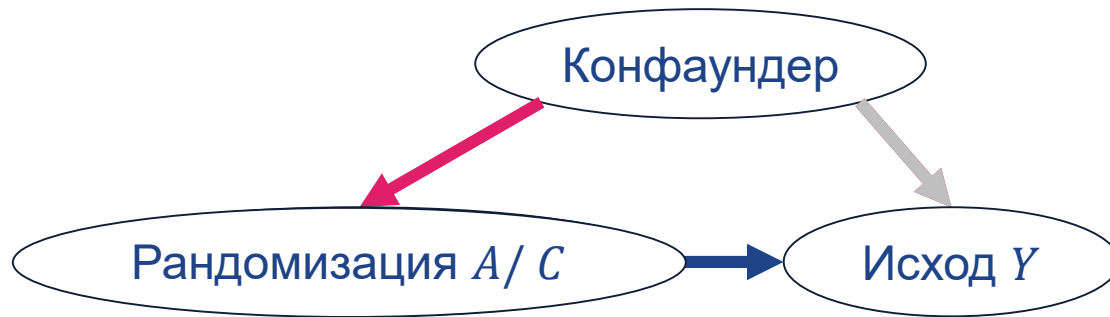
$$\Delta Y = Y(X = A) - Y(X = C)$$

Вывод для
популяции

РАНДОМИЗАЦИЯ И ЕЕ МИШЕНЬ

С точки зрения причинно-следственного вывода

- **Рандомизация воздействия** – процесс принятия решения о том, какой пациент какое воздействие получит, с использованием элемента случайности – такого, что вероятность попадания пациентов в каждую группу воздействия одинакова (для всех пациентов)
- **Позволяет избавиться от одного источника систематических ошибок – конфаундинга, влияющего на назначение терапии (confounding by indication)**



Кроссовер:

$$\text{Эффект воздействия} = \overline{\Delta Y}$$



Параллельный дизайн:
Эффект воздействия –

~~$$\bar{Y}(X = A) - \bar{Y}(X = C)$$~~

РАНДОМИЗАЦИЯ НЕ ДЛЯ ТОГО, ЧТОБЫ...

И нет, сама по себе она не обязана...

- **Сбалансировать группы по исходным характеристикам**

→ Это вообще не нужно в РКИ

- ☑ Рандомизация балансирует группы по потенциальным исходам

- **Защитить от других систематических ошибок**, помимо confounding by indication

→ Для каждого источника систематических ошибок – свой способ их предотвращения, коррекции или анализа чувствительности

- ☑ **Интеркуррентные события в концепции эстиманда (ICH E9 Addendum)**

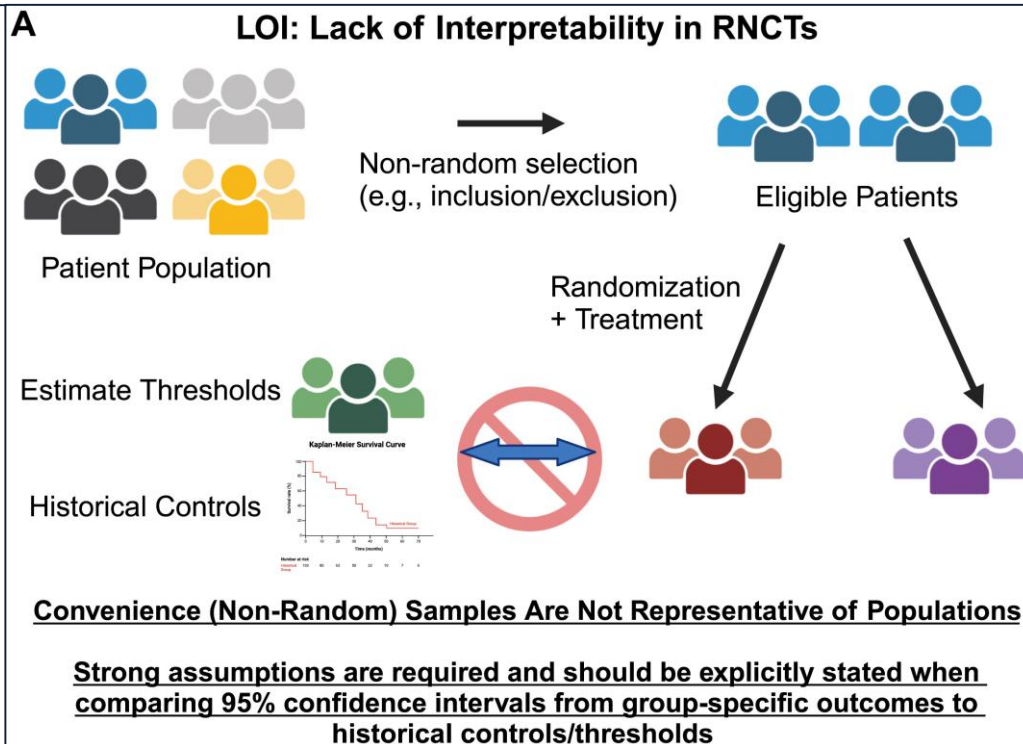
РАНДОМИЗАЦИЯ НЕ ДЛЯ ТОГО, ЧТОБЫ...

И нет, сама по себе она не обязана...

- **Дать возможность делать выводы** о каком-либо параметре распределения какого-либо показателя в целевой популяции **в группе** воздействия или в контрольной группе (например, о медиане выживаемости на терапии *B*)
 - Такую возможность может дать только **случайная выборка пациентов из целевой популяции**
 - ☑ При этом рандомизация воздействия позволяет делать **выводы о разнице между группами (разнице в средних или отношениях – HR, OR, RR, IRR)** при отсутствии модификаторов эффекта / их распределении, как в целевой популяции / их включении в модель

A meta-epidemiological analysis of post-hoc comparisons and primary endpoint interpretability among randomized noncomparative trials in clinical medicine

Alexander D. Sherry^a · Pavlos Msaouel^{b,c} · Ethan B. Ludmir^{d,e}



Clinical trials, which require patient consent and use inclusion and exclusion criteria to select patients, study convenience samples rather than random samples and therefore cannot be used to reliably estimate population outcomes

РАНДОМИЗАЦИЯ НЕ ДЛЯ ТОГО, ЧТОБЫ...

И нет, сама по себе она не обязана...

- **Дать точную оценку эффекта**
(узкий доверительный интервал)
 - Это функция того, что может **уменьшить** (узкие критерии отбора) или **«объяснить»** **вариабельность исходов внутри групп**
 - ☑ Стратификация + блоковая рандомизация с последующим стратифицированным анализом
 - ☑ Регрессия с ковариатами для прогностических признаков

ЗАЧЕМ НУЖНА РАНДОМИЗАЦИЯ

Второго мнения быть не может

- Чтобы мы могли сделать **валидный** причинно-следственный вывод об эффекте интересующего нас воздействия на интересующий нас исход (= корректно оценить его эффект **по сравнению с** контрольной группой) в целевой популяции

ПОЧЕМУ RNCT ОКСЮМОРОН

Рандомизация не работает «офф-лейбл»

- Рандомизация подразумевает дальнейшее сравнение групп
- Если целью исследования не является сравнение групп рандомизации в RNCT, то рандомизация по ним не нужна
- По оценкам средних значений исходов (выживаемость, частота ответа и т.п.) в каждой группе RNCT нельзя будет сделать вывод о них в целевой популяции из-за неслучайности выборки участников исследования из нее
- Сравнение средних значений исходов в отдельном рукаве RNCT с пороговым значением не позволяет сделать вывод об эффективности соответствующей терапии (это инструмент для решения описательной, а не каузальной задачи)
- Сравнение средних значений исходов в отдельном рукаве RNCT с исторической или параллельной внешней контрольной группой требует коррекции конфаундинга

ПРИЧИНЫ РАСПРОСТРАНЕНИЯ RNCT*

Гипотеза 4 («наболело»)

- **Отсутствие диалога между клиницистами и статистиками**
- Нехватка квалифицированных статистиков
- Низкий уровень образования в области статистики и ее философии (!)
- Большая вариация квалификации среди «статистиков»
- Публикация в высокорейтинговом журнале как мерило успеха и пример для подражания
- Восприятие опровержений от статистиков как части «научной дискуссии», обмена «мнениями» двух сторон, каждое из которых должно быть услышано аудиторией
- Закон Брандолини: количество энергии, необходимой для опровержения чуши, на порядок больше, чем требуется для её производства
- Асимметрия ролей (почему статистики не пишут в своих журналах статьи по онкологии??)
- Сведение роли статистика к расчету p-значений и объема выборки

*** и многих других статистических оксюморонов**

Что делать?



f2harrell 

Jun 1

I'd like to better understand the original motive for such a stupid design. Perhaps it had to be the perception that it was lowering the bar, to make finding treatment benefit easier.



simongates



 Pavlos_Msaouel

This is all fascinating. The message that I take from it is that we can all get it very wrong sometimes. So probably a good idea to be open to criticism and prepared to learn.

УЧИТЬСЯ, УЧИТЬСЯ И ЕЩЕ РАЗ УЧИТЬСЯ

Предложение 1 – очевидное

- **Налаживать диалог:**
 - Статистика без доменного знания не имеет смысла
 - Статистика начинается с дизайна исследования
 - Статистический вывод \neq исследовательский вывод
 - Научные исследования \neq принятие решений
- **Статистика – наука о неопределенности**, а не о принятии решений: доверительные интервалы гораздо более информативны, чем р-значения
 - Не ждите / не ищите простого ответа на сложный вопрос
- **Изучение статистики начинается с эпидемиологии**
- На пути от постановки исследовательского вопроса до получения ответа на него делаются **явные и неявные допущения** – для полноценного исследовательского вывода с учетом ограничений необходимо их понимать



To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of.

— Ronald Fisher —

РКИ: «ПОСЧИТАЙТЕ ОБЪЕМ ВЫБОРКИ ПОМЕНЬШЕ»

Предложение 2 – неочевидное

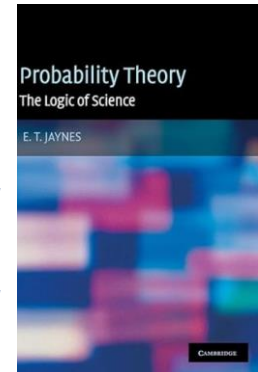
- Стратификация / регрессия с ковариатами для прогностических признаков
 - Если можно не категоризовать ковариаты, не нужно категоризовать
- Последовательные дизайны с мягкими границами для остановки по бесполезности
- Перекрестные РКИ, N-of-1?
- Суррогатные конечные точки?

ЕСТЬ ДИЗАЙНЫ ПОМИМО РКИ И RNCT

Предложение 3 – логичное

- Если задача – сравнение с определенным компаратором при невозможности провести РКИ, то **нерандомизированное сравнительное исследование, в том числе одорукавное исследование с внешней контрольной группой**
 - Для дизайна рекомендуется погрузиться в изучение теории причинно-следственного вывода и эмуляции целевого исследования
 - Не «псевдорандомизацией»* единой: у propensity score matching (PSM) есть масса недостатков, но есть и масса более приемлемых альтернатив

Whenever there is a randomized way of doing something, then there is a nonrandomized way that delivers better performance but requires more thought



* с точки зрения статистики, это некорректный термин

RNST – НИКОГДА-НИКОГДА?

...



СПАСИБО ЗА ВНИМАНИЕ!

11-13 ДЕКАБРЯ 2025 · МОСКВА

XXIX РОССИЙСКИЙ ОНКОЛОГИЧЕСКИЙ КОНГРЕСС 2025

