

МЕТОДЫ ИССЛЕДОВАНИЯ И СОСТАВЛЕНИЕ СТАТЕЙ. CONSORT 2010 — ОБЪЯСНЕНИЕ И УТОЧНЕНИЕ: ОБНОВЛЕННЫЕ РЕКОМЕНДАЦИИ ПО СОСТАВЛЕНИЮ СТАТЬИ О РАНДОМИЗИРОВАННЫХ КЛИНИЧЕСКИХ ИССЛЕДОВАНИЯХ С ПАРАЛЛЕЛЬНЫМИ ГРУППАМИ — ЧАСТЬ I

Оригинал опубликован в *BMJ* 2010;340:c869. DOI: 10.1136/bmj.c869

Перевод подготовлен: Эсенбекова Э.Э., Гришина М.С., Ерусланова К.А.

Резюме

В настоящее время существуют неопровержимые доказательства, свидетельствующие о недостаточном качестве статей о рандомизированных контролируемых исследованиях (РКИ). При отсутствии понятных для читателей статей сложно судить о надежности и достоверности результатов исследований, трудно извлекать информацию для систематических обзоров. Недавние методологические анализы указывают на то, что недостатки при составлении статьи и дизайна связаны с систематическими ошибками при оценке эффективности лечения. Такие ошибки наносят серьезный ущерб РКИ, которые считаются золотым стандартом для оценки вмешательств из-за их способности минимизировать или избежать необъективности.

Группой ученых и редакторов разработано заявление CONSORT (Консолидированные (общие) стандарты предоставления результатов об исследованиях) для улучшения качества представления информации о РКИ. Многие ведущие медицинские журналы и крупные редакционные группы одобрили стандарты CONSORT. Руководство CONSORT облегчает критическую оценку и интерпретацию РКИ.

В этом номере мы публикуем первую часть перевода документа CONSORT.

Ключевые слова: рандомизированное клиническое исследование; протокол; статья; объем выборки; статистика.

Для цитирования: Методы исследования и составление статей. CONSORT 2010 — объяснение и уточнение: обновленные рекомендации по составлению статьи о рандомизированных клинических исследованиях с параллельными группами. Часть I. *Проблемы геронауки*. 2024; 1(5): 30–54.

RESEARCH METHODS & REPORTING. CONSORT 2010 — EXPLANATION AND ELABORATION: UPDATED GUIDELINES FOR REPORTING PARALLEL GROUP RANDOMISED TRIALS — PART I

Original article published in *BMJ* 2010;340:c869. DOI: 10.1136/bmj.c869

Translation prepared by: Esenbekova E.E., Grishina M.S., Eruslanova K.A.

Abstract

Overwhelming evidence shows the quality of reporting of randomised controlled trials (RCTs) is not optimal. Without transparent reporting, readers cannot judge the reliability and validity of trial findings nor extract information for systematic reviews. Recent methodological analyses indicate that inadequate reporting and design are associated with biased estimates of treatment effects. Such systematic error is seriously damaging to RCTs, which are considered the gold standard for evaluating interventions because of their ability to minimise or avoid bias. A group of scientists and editors developed the CONSORT (Consolidated Standards of Reporting Trials) statement to improve the quality of reporting of RCTs. Many leading medical journals and major international editorial groups have endorsed the CONSORT statement. The CONSORT statement facilitates critical appraisal and interpretation of RCTs. In this issue we publish the first part of the translation of the CONSORT document.

Keywords: randomized clinical trial; protocol; article; sample size; statistics.

For citation: Research methods & reporting. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. Part I. *Problems of Geroscience*. 2024; 1(5): 30–54.

«Вся медицина зависит от прозрачности статей клинических исследований»

Хорошо разработанные и должным образом проведенные клинические исследования предоставляют наиболее достоверные доказательства эффективности медицинских вмешательств, но исследования с неадекватными методами связаны с систематической ошибкой, особенно в преувеличении оценки эффектов лечения. Субъективные результаты плохо разработанных и зарегистрированных исследований могут привести к принятию ошибочных решений на всех уровнях оказания медицинской помощи, от лечения конкретного пациента до формирования национальной политики общественного здравоохранения.

Критическая оценка качества клинических исследований возможна в том случае, если в опубликованных статьях представлены точные и подробные описания дизайна, выполнение и анализ данных РКИ. Однако такие отчеты, отнюдь не являясь прозрачными, часто представляют неполную информацию о РКИ [6–9], что усугубляет проблемы, связанные с плохой методологией [10–15].

НЕПОЛНОЕ И НЕТОЧНОЕ ПРЕДСТАВЛЕНИЕ РЕЗУЛЬТАТОВ В ОТЧЕТАХ

Во многих обзорах отмечены недостатки статей о результатах клинических исследований. Например, информация о методе, использованном в исследовании для распределения участников по группам

сравнения, была представлена только у 21% из 519 исследований, указанных в PubMed в 2000 году [16], и только у 34% из 616 — в 2006 году [17]. Также только 45% клинических исследований, указанных в PubMed в 2000 году [16], и 53%, указанных в 2006 году [17], определили первичную и конечную точку, и только 27% в 2000 и 45% в 2006 году сообщили о расчете размера выборки. Отчетность о РКИ не только часто бывает неполной, но также иногда неточной. Из 119 статей, в которых говорится, что все участники были включены в окончательный анализ в составе тех групп, в которые они были включены изначально, 15 (13%) исключали пациентов или не анализировали всех пациентов, как это было указано [18]. Во многих других обзорах было обнаружено, что недостатки распространены в статьях специализированных журналов [16, 19], а также в журналах, изданных на других языках, кроме английского [20, 21].

Правильная рандомизация снижает погрешность отбора при включении в исследование и является важнейшим компонентом высокого качества РКИ [22]. Успешная рандомизация зависит от двух этапов: 1) генерация непредсказуемой последовательности распределения участников и 2) сокрытие этой последовательности от исследователей, регистрирующих участников (вставка 1) [2, 23].

К сожалению, несмотря на важную роль, информация о методах, используемых для распределения участников исследования по группам вмешательства, как правило, представляется недостаточно. Например, 5% из 206 статей, опубликованных в журналах по акушерству и гинекологии как

рандомизированные контролируемые исследования, не были по-настоящему рандомизированы [23]. Эта оценка является заниженной, поскольку большинство статей в настоящее время не содержат адекватную информацию о методах распределения участников исследования [20, 23, 30–33].

УЛУЧШЕНИЕ ПРЕДОСТАВЛЕНИЯ РЕЗУЛЬТАТОВ РАНДОМИЗИРОВАННЫХ КОНТРОЛИРУЕМЫХ ИССЛЕДОВАНИЙ: ЗАЯВЛЕНИЕ CONSORT

Ребекка ДерСимонян и ее коллеги предположили, что «редакторы журналов могли бы значительно улучшить статьи о клинических исследованиях, предоставив авторам перечень пунктов, которые должны строго выполняться» [34]. В начале 1990-х две группы редакторов журналов, исследователей и методологов независимо друг от друга опубликовали рекомендации по составлению отчетов об исследованиях [35, 36]. В последующей редакционной статье Ренни призвал две группы встретиться и разработать общие рекомендации [37]; результатом стало заявление CONSORT (Консолидированные (единые/общие) стандарты представления отчетов об исследованиях) [38].

Заявление CONSORT (в дальнейшем CONSORT) содержит контрольный список основных вопросов, которые должны быть включены в статьи о результатах РКИ, и схему документирования потока участников в ходе исследования (схема проведения исследования). CONSORT предназначен для статей о РКИ с двумя группами параллельных дизайнов. Большая часть CONSORT также применима к более широкому классу дизайнов исследований, таких как исследования эквивалентности, факториальные, кластерные и перекрестные исследования. Были опубликованы дополнения к контрольному списку CONSORT для статей об исследованиях с некоторыми из этих дизайнов [39–41], а также для статей определенных типов данных (вред [42]), типов вмешательств (немедикаментозное лечение [43], нетрадиционные вмешательства [44]) и аннотаций [45].

Цель CONSORT — предоставить авторам рекомендации о том, как улучшить предоставление результатов об их исследованиях. Статьи об исследованиях должны быть четкими, полными и прозрачными. Читатели, рецензенты и редакторы журналов также могут использовать CONSORT, который может помочь им критически оценивать и интерпретировать статьи РКИ. Однако CONSORT не предназначается для использования в качестве инструмента оценки качества исследований. Скорее содержание CONSORT сосредоточено на элементах, связанных с внутренней и внешней валидностью испытаний. Многие элементы, четко не упомянутые в CONSORT, также должны быть включены в статью, например информация об одобрении исследования комитетом по этике, получение информированного согласия от участников и, где это уместно, о существовании комитета

Вставка 1. Распределение участников исследования по группам лечения. Рандомизация: принципы и особенности

Метод, используемый для распределения участников исследования по группам вмешательств, является важным аспектом дизайна клинического исследования. Случайное распределение, которое регулярно успешно используется в исследованиях на протяжении более 50 лет [24], является предпочтительным методом. Рандомизация имеет три основных преимущества [25]. Во-первых, при правильном применении устраняет систематическую ошибку отбора, уравнивая известные и неизвестные прогностические факторы при распределении в группы лечения. Без рандомизации сравнение лечебных вмешательств может оказаться предвзятым, сознательно или нет, из-за назначения определенным участникам определенного типа лечения. Во-вторых, случайное распределение позволяет использовать теорию вероятностей для выражения степени правдоподобия того, что любая разница в результатах между группами вмешательств просто отражает случайность [26]. В-третьих, случайное распределение в некоторых ситуациях способствует тому, чтобы исследователи, участники и эксперты, оценивающие исходы, не знали об идентичности лечения, возможно, за счет использования плацебо, что снижает вероятность возникновения систематических ошибок после назначения лечения [27]. Из этих трех преимуществ снижение систематических ошибок при включении участников в исследование обычно является наиболее важным [28].

Успешная рандомизация на практике зависит от двух взаимосвязанных аспектов — адекватного создания непредсказуемой последовательности распределения участников в группы исследования и сокрытия этой последовательности до тех пор, пока не произойдет назначение вмешательства [2, 23]. Ключевой вопрос заключается в том, известна ли или предсказуема эта система распределения для лиц, участвующих в отнесении участников к той или иной группе сравнения [29]. Таким образом, система распределения лечения должна быть настроена так, чтобы человек, проводящий включение пациентов, не знал заранее о том, какое лечение получит следующий участник. Этот процесс называется сокрытием распределения [2, 23]. Надлежащее сокрытие распределения участников в группы вмешательства является фактором, способствующим сокрытию информации о предстоящих назначениях пациентам, тогда как эффективная случайная последовательность предотвращает безошибочное прогнозирование будущих назначений, основываясь на информации о предыдущих назначениях.

по безопасности и мониторингу данных. Кроме того, должны быть представлены надлежащим образом и другие аспекты испытаний, такие как информация, необходимая для анализа эффективности затрат и оценки качества жизни [46–48].

С момента публикации в 1996 году CONSORT поддерживали более 400 журналов (www.consort-statement.org) и несколько редакционных групп, например Международный комитет редакторов медицинских журналов [49]. Внедрение CONSORT в журналы улучшило качество статей о результатах РКИ [17, 50, 51]. Тем не менее CONSORT является развивающейся инициативой, и заявление CONSORT периодически

пересматривается [3]. Последний раз CONSORT пересматривался девять лет назад, в 2001 году [52–54]. С тех пор доказательная база для информирования CONSORT значительно расширилась; эмпирические данные высветили новые проблемы, связанные с отчетностью РКИ, такие как выборочная отчетность о результатах [55–57]. Поэтому в январе 2007 года в Канаде было созвано совещание CONSORT-группы для пересмотра заявления CONSORT 2001 года и сопровождающего его пояснительного и уточняющего документа. Пересмотренный контрольный перечень показан в таблице 1, а непересмотренная блок-схема — на рисунке 1 [52–54].

Таблица 1.

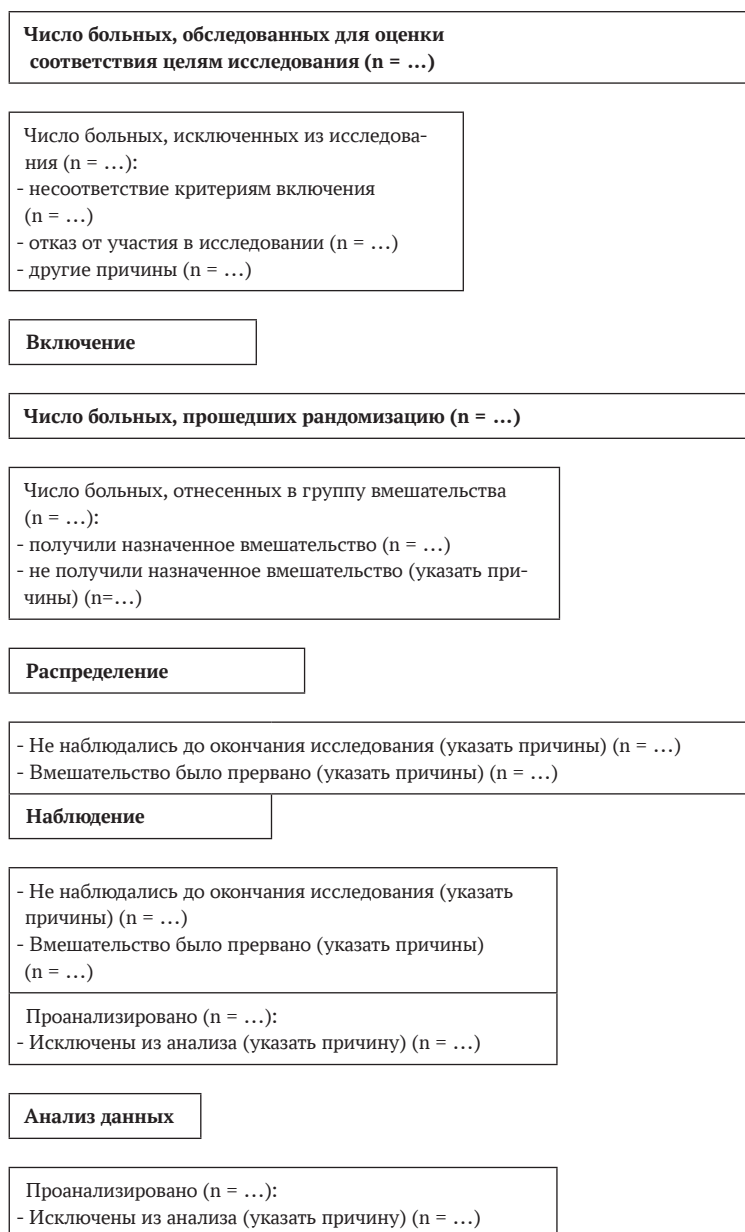
CONSORT 2010 — контрольный список информации, который следует учитывать при составлении статьи о рандомизированном исследовании

Раздел/тема	Пункт/Номер	Контрольная точка
Название и абстракт		
	1a	Идентификация как рандомизированное исследование в названии
	1b	Структурированное резюме (абстракт) дизайна исследования, методов, результатов и выводов (конкретные рекомендации смотрите в CONSORT для тезисов)
Введение		
Обоснование и цели	2a	Научные предпосылки/ актуальность/ основа и обоснование/ рационализация проведения исследования
	2b	Конкретные цели или гипотезы
Методы		
Дизайн исследования	3a	Описание дизайна исследования (например, параллельное, факторное), включая коэффициент распределения
	3b	Важные изменения, внесенные в методы после начала исследования (такие как критерии включения) с указанием причин
Участники	4a	Критерии включения для участников
	4b	Условия и местоположение, где производился сбор данных
Вмешательства	5	Описание вмешательств для каждой группы с приведением деталей, достаточных для воспроизведения, включая то, как и когда они фактически применялись
Выводы (результаты)	6a	Полностью определенные, предварительно конкретизированные первичные и вторичные показатели выводов, включая указание на то, как и когда они были оценены
	6b	Любые изменения результатов исследования после его начала с указанием причин
Размер выборки	7a	Определение размера выборки
	7b	При необходимости объяснение любого промежуточного анализа и оснований для прекращения исследования
Рандомизация		
Генерация последовательности	8a	Метод, используемый для генерации последовательности случайного распределения участников исследования
	8b	Тип рандомизации, детализация любых ограничений (таких как объединение в группы и размер групп)

Раздел/тема	Пункт/Номер	Контрольная точка
Механизм сокрытия распределения участников в группы исследования	9	Механизм, используемый для осуществления случайной последовательности распределения участников в группы исследования (такой как последовательно пронумерованные контейнеры), описывающий любые шаги, предпринятые для сокрытия этой последовательности до тех пор, пока не будут назначены вмешательства
Реализация	10	Разработка случайной последовательности распределения участников в группы исследования, регистрация участников, распределение участников по группам вмешательств
Ослепление	11a	Если было проведено слепое исследование, указать, кто был ослеплен после определения в группы вмешательств (например, участники, исполнители, лица, оценивающие исходы) и каким образом это проводилось
	11b	Если применимо, описание сходства вмешательств
Статистические методы	12a	Статистические методы, используемые для сравнения групп по первичным и вторичным результатам
	12b	Методы для дополнительного анализа, такие как анализ подгрупп и скорректированный анализ
Результаты		
Поток участников (настоятельно рекомендуется использовать диаграмму)	13a	Количество участников при случайном распределении на каждом этапе исследования, получивших предполагаемое лечение; оценка первичных результатов
	13b	Для каждой группы указаны потери и исключения участников после рандомизации с причинами
Набор участников исследования	14a	Сроки, определяющие периоды набора участников и последующего наблюдения
	14b	Причины прекращения или перерыва исследования
Исходные данные	15	Таблица, показывающая исходные демографические и клинические характеристики для каждой группы
Анализ данных	16	Число участников в каждой из групп, данные о которых были включены в анализ любого вида (в знаменателе), а также информация о том, проводился ли анализ
Результаты и оценка	17a	Для каждого первичного и вторичного исхода результаты представлены по каждой группе, а также предполагаемый размер эффекта и его точность (например, доверительный интервал 95%)
	17b	Для бинарных результатов рекомендуется представление как абсолютных, так и относительных величин эффекта
Дополнительный анализ	18	Результаты любых других проведенных анализов, включая анализ подгрупп и скорректированный анализ, отличающийся от предварительно определенного анализа
Вред	19	Указать все важные вредные или непреднамеренные последствия в каждой группе (для получения конкретных указаний см. CONSORT для вреда)
Обсуждение		
Ограничения	20	Ограничения исследования, устранение источников потенциальной систематической ошибки, неточности и при необходимости множественности анализов
Обобщение	21	Обобщение (внешняя валидность, применимость) результатов исследования
Интерпретация	22	Интерпретация в соответствии с результатами, сопоставление преимуществ и недостатков, рассмотрение других значимых факторов
Другая информация		
Регистрация	23	Регистрационный номер и название исследования в реестре
Протокол	24	Информация о получении доступа к полному протоколу исследования, при наличии
Финансирование	25	Источники финансирования и другая поддержка (например, обеспечение лекарственными средствами), роль финансирующих организаций

Рисунок 1.

**Блок-схема проведения этапов параллельного рандомизированного исследования двух групп
(набор участников исследования, место проведения вмешательства, последующее наблюдение
и анализ данных)**



ЗАЯВЛЕНИЕ CONSORT 2010: РАЗЪЯСНЕНИЕ И ДОРАБОТКА

В ходе пересмотра CONSORT 2001 года стало ясно, что разъяснение и уточнение основных принципов заявления CONSORT поможет исследователям и другим лицам при написании или оценке статей об исследованиях. Статья CONSORT с объяснениями и уточнениями [58] была опубликована в 2001 году вместе с версией заявления CONSORT 2001 года, в котором обсуждались логическое обоснование и научная основа каждого пункта и приводились опубликованные примеры хорошей отчетности. Обоснование пересмотра этой

статьи аналогично обоснованию пересмотра заявления, описанного выше. Ниже мы кратко опишем основные дополнения и сокращения в этой версии статьи с объяснениями и доработками.

РАЗЪЯСНЕНИЕ И ДОРАБОТКА CONSORT 2010: ИЗМЕНЕНИЯ

Мы внесли несколько изменений по существу и улучшающих текст поправок в эту версию пояснительного документа CONSORT (полная информация приведена в версии заявления CONSORT от 2010 года [59]). Некоторые из них отражают изменения

в контрольном списке CONSORT; в контрольном списке CONSORT 2010 есть три новых пункта, например пункт 24, в котором авторам предлагается сообщить, где можно получить доступ к протоколу их исследований. Мы также обновили некоторые существующие объяснения, в том числе добавили более свежие ссылки на методологические данные и использовали несколько лучших примеров. Мы удалили глоссарий, который теперь доступен на веб-сайте CONSORT (www.consort-statement.org). Там, где это возможно, мы описываем результаты соответствующих эмпирических исследований. Многие превосходные книги по клиническим исследованиям предлагают более полное обсуждение методологических вопросов [60–62]. Наконец, для удобства мы иногда используем термины “лечение” и “пациенты”, хотя признаем, что не все вмешательства, оцениваемые в РКИ, являются методами лечения и не все участники являются пациентами.

Контрольный список (чек-лист)

Название и абстракт

Пункт 1а. Идентификация, рандомизированное исследование в названии

Пример: «Борьба с курением с помощью пероральных никотиновых ингаляторов: двойное слепое рандомизированное клиническое исследование с оценкой эффективности и безопасности» [63].

Объяснение. Возможность найти соответствующую публикацию в электронной базе данных зависит в значительной степени от того, как она была проиндексирована. Статья может быть не отнесена к рандомизированному исследованию, если авторы не сообщают об этом явным образом [64]. Для обеспечения надлежащей индексации исследования как РКИ авторы должны четко указать в заголовке своего сообщения, что участники исследования были распределены в группы сравнения рандомизированно.

Пункт 1б. Структурированное резюме (абстракт) дизайна исследования, методов, результатов и выводов

Конкретные рекомендации смотрите в CONSORT для тезисов [45, 65].

Объяснение. Четкие, прозрачные и достаточно подробные резюме важны, поскольку читатели часто основывают свою оценку исследования на такой информации. Некоторые читатели используют резюме в качестве инструмента проверки, чтобы решить, стоит ли читать статью полностью. Поскольку не все исследования находятся в свободном доступе, а некоторые медицинские работники не имеют доступа к полным статьям, медицинские решения иногда принимаются на основе резюме рандомизированных исследований [66].

Резюме к журналу должно содержать достаточно информации об исследовании, чтобы служить точным протоколом о его проведении и результатах, предоставляя оптимальную информацию об исследовании в рамках ограничений по объему и формату журнала. Правильно составленное и написанное резюме

помогает читателям быстро оценить релевантность результатов и найти соответствующие статьи в электронных базах данных [67]. Резюме должно точно отражать то, что включено в полную журнальную статью, и не должно включать информацию, которая отсутствует в основном тексте. Исследования, сравнивающие точность информации, представленной в резюме журнала, с информацией, представленной в тексте полной публикации, обнаружили утверждения, которые не соответствуют тексту или отсутствуют в полной статье [68–71]. И наоборот, исключение важного пункта из резюме может серьезно исказить интерпретацию результатов исследования [42, 72].

Недавнее расширенное заявление CONSORT предоставляет список основных элементов, которые авторы должны включать при сообщении основных результатов рандомизированного исследования в резюме журнала (или конференции) (табл. 2) [45]. Мы настоятельно рекомендуем использовать структурированные резюме для статей о рандомизированных исследованиях. Они предоставляют читателям информацию об исследовании под рядом заголовков, касающихся дизайна, проведения, анализа и интерпретации [73]. Некоторые исследования показали, что структурированные резюме чаще более высокого качества, чем более традиционные описательные резюме [74, 75], а также они позволяют читателям легче находить информацию [76]. Мы понимаем, что многие журналы разработали свою собственную структуру и лимит слов для резюме. В наши намерения не входит предлагать изменения в этих форматах, но мы даем рекомендации о том, какую информацию необходимо сообщать в резюме.

Введение

Пункт 2а. Научные предпосылки/ актуальность/ основа и обоснование/ рационализация проведения исследования

Пример: «Операция является методом выбора для пациентов с немелкоклеточным раком легкого (НМРЛ) стадии I и II. Метаанализ объединил результаты восьми рандомизированных исследований, где сравнивалось применение хирургических методов лечения НМРЛ с хирургическим лечением в сочетании с химиотерапией на основе цисплатина. По результатам выявлено незначительное ($p = 0,08$) улучшение выживаемости примерно на 5% за 5 лет (с 50% до 55%) при применении второй тактики. На момент разработки настоящего исследования (середина 1990-х) адъювантная химиотерапия еще не стала стандартной клинической практикой. Клиническое обоснование неоадъювантной химиотерапии состоит из трех частей: можно добиться регрессии первичного рака, тем самым облегчая и упрощая лечение, или сокращения последующей операции; невыявленные микрометастазы можно было устранить в начале лечения; может иметь место ингибирование предполагаемого стимула к остаточному раку факторами роста, высвобождаемыми хирургическим вмешательством

Таблица 2.

**Элементы, которые необходимо включать в резюме (абстракт) журнала при сообщении
о рандомизированном исследовании**

Пункт	Описание/характеристика
Авторы	Контактная информация
Дизайн исследования	Описание дизайна исследования (например, параллельного, кластерного, исследования не меньшей эффективности (безопасности))
Методы:	
Участники	Критерии включения для участников и условия, где производился сбор данных
Вмешательства	Вмешательства, предназначенные для каждой группы
Цель	Конкретные цели или гипотезы
Выводы	Полностью определенные основные выводы данной статьи
Рандомизация	Порядок распределения участников по группам вмешательств
Ослепление	Были ли участники, медицинские работники и лица, оценивающие исходы, ослеплены после определения в группы вмешательств
Результаты:	
Число рандомизированных	Число участников, распределенных по группам вмешательств
Набор участников исследования	Состояние исследования
Проанализированные цифры	Число участников, проанализированных в каждой группе вмешательств
Результаты	Для первичного исхода, результаты, представленные по каждой группе, а также предполагаемый размер эффекта и его точность
Вред	Важные неблагоприятные события или побочные эффекты
Выводы	Общая интерпретация результатов
Регистрация исследования	Регистрационный номер и название реестра исследований
Финансирование	Источники финансирования

с последующим заживлением раны. Таким образом, настоящее исследование было организовано для сравнения у пациентов с резектабельным НМРЛ только хирургического вмешательства с тремя циклами химиотерапии на основе препаратов платины с последующим хирургическим вмешательством с точки зрения общей выживаемости, качества жизни, патологической стадии, показателей резектабельности, степени хирургического вмешательства, а также времени и места рецидива» [77].

Объяснение. Как правило, раздел «Введение» состоит из текста, изложенного в свободной форме, в котором авторы объясняют научную основу и обоснование своего исследования, а также научное обоснование его целесообразности. В данный раздел также целесообразно включить цели исследования (см. пункт 2b). Обоснование может быть пояснительным (например, для оценки возможного влияния лекарственного средства на функцию почек) или прагматическим (например, разработка клинических рекомендаций к практическому применению путем сопоставления клинических эффектов двух альтернативных методов лечения). Авторы должны предоставить все имеющиеся доказательства преимуществ и недостатков активных вмешательств, включенных в исследование, а также должны

предлагать правдоподобное объяснение возможного механизма действия исследуемого вмешательства, особенно если ранее оно применялось редко или вообще не применялось [78].

В Хельсинкской декларации говорится, что биомедицинские исследования с участием людей должны основываться на глубоком знании научной литературы [79]. Это означает, что неэтично подвергать людей неоправданному риску при проведении исследований. Было показано, что в некоторых клинических исследованиях нет необходимости, потому что вопрос, который они затрагивали, был или мог быть решен путем систематического обзора существующей литературы [80, 81]. Таким образом, в введении должна быть обоснована необходимость проведения нового исследования. В идеале введение должно содержать ссылку на систематический обзор предыдущих подобных исследований или примечание на отсутствие таких исследований [82].

Пункт 26. Конкретные цели или гипотезы

Пример: «В текущем исследовании мы проверили гипотезу о том, что политика активного ведения первородящих: 1) снизит частоту кесарева сечения; 2) снизит частоту затянувшихся родов; 3) не повлияет на удовлетворенность матери опытом родов» [83].

Объяснение. Цели — это вопросы, на которые должно было ответить исследование. Они часто связаны с эффективностью конкретного терапевтического или профилактического вмешательства. Гипотезы — это заранее заданные вопросы, ответы на которые проверяются при достижении целей. Гипотезы более конкретны, чем цели, и поддаются явной статистической оценке. На практике цели и гипотезы не всегда легко отличить друг от друга. В большинстве статей о РКИ содержится адекватная информация о целях исследований и гипотезах [84].

Методы

Пункт 3а. Описание дизайна исследования (например, параллельного, факторного), включая коэффициент распределения

Пример: «Это было многоцентровое, стратифицированное (от 6 до 11 лет и от 12 до 17 лет, с несбалансированной рандомизацией [2:1]), двойное слепое, плацебо-контролируемое исследование с параллельными группами, проведенное в Соединенных Штатах (41 участок)» [85].

Объяснение. Слово «дизайн» часто используется для обозначения всех аспектов организации исследования, но оно также имеет более узкое толкование. Многие конкретные аспекты более широкого дизайна исследования, включая детали рандомизации и ослепления, рассматриваются в других разделах контрольного списка CONSORT. Здесь мы ищем информацию о типе исследования, таком как параллельная группа или факториальное исследование, и концептуальной основе, такой как превосходство или неполноценность, и других связанных вопросах, не рассмотренных в другом месте контрольного списка.

Заявление CONSORT сосредоточено в основном на исследованиях с участниками, индивидуально рандомизированными в одну из двух параллельных групп. На самом деле чуть больше половины опубликованных исследований имеют такой дизайн [16]. Основными альтернативными планами являются параллельный, перекрестный, кластерный и факторный дизайны [40]. Кроме того, большинство исследований нацелено на выявление превосходства нового вмешательства, если оно существует, но другие предназначены для оценки не меньшей эффективности или эквивалентности [39]. Важно, чтобы исследователи четко описывали эти аспекты своего исследования, включая единицу рандомизации (например, пациент, врач общей практики, поражение). Желательно также включить эти детали в резюме (см. пункт 1b).

Если используется менее распространенный дизайн, авторам рекомендуется объяснить свой выбор, особенно потому, что такой дизайн может подразумевать необходимость большего размера выборки или более сложного анализа и интерпретации. Однако в большинстве исследований используется одинаковая рандомизация (например, 1:1 для двух групп), также полезно указать коэффициент распределения в явном виде. Для исследований лекарственных средств

также может иметь значение указание фазы исследования (I-IV).

Пункт 3б. Важные изменения, внесенные в методы после начала исследования (такие как критерии включения) с указанием причин

Пример: «Пациенты были случайным образом распределены в одну из шести параллельных групп, первоначально в соотношении 1:1:1:1:1:1, для получения одного из пяти режимов отамиксабана. Проводился активный контроль нефракционированного гепарина... Независимый комитет по мониторингу данных проанализировал открытые данные на предмет клинической эффективности и безопасности отамиксабана у пациентов; никаких промежуточных анализов эффективности или бесполезности не проводилось. Во время исследования комитет рекомендовал, чтобы группа, получавшая самую низкую дозу отамиксабана (0,035 мг/кг/ч), была исключена из исследования из-за клинических признаков неадекватной антикоагулянтной терапии. Протокол был немедленно изменен в соответствии с этой рекомендацией, и впоследствии участники были случайным образом распределены в соотношении 2:2:2:2:1 к оставшимся группам отамиксабана и контрольной группе соответственно» [86].

Объяснение. Некоторые исследования могут начинаться без какого-либо фиксированного плана (т.е. быть полностью исследовательскими), но большинство из них имеют протокол, подробно описывающий, как будет проводиться исследование. Допустимы отклонения от первоначального протокола, поскольку нельзя предсказать все возможные изменения обстоятельств в ходе исследования. Таким образом, некоторые исследования могут иметь важные изменения в методах после их начала.

Изменения могут быть связаны с получением внешней информации из других исследований или внутренними финансовыми трудностями, а также зависеть от недостаточно быстрого накопления выборки. Такие изменения протокола следует вносить, не нарушая анонимности накопленных данных об исходах участников. В некоторых исследованиях в компетенцию независимого комитета по мониторингу данных входит возможность рекомендовать изменения протокола на основе просмотра открытых данных. Такие изменения могут повлиять на методы исследования (например, изменения в схемах лечения, критериях приемлемости, коэффициенте рандомизации или продолжительности наблюдения) или на проведение исследования (например, отказ от работы с центром с низким качеством данных) [87].

Некоторые исследования имеют формальный «адаптивный» дизайн. Не существует общепринятого определения этих планов, но рабочим определением может быть «многоэтапный план исследования, в котором используется накопление данных для принятия решения о том, как изменить аспекты исследования без ущерба для достоверности и целостности

исследования» [88]. Модификации обычно зависят от размера выборки и количества групп лечения и могут привести к более быстрому принятию решений и более эффективному использованию ресурсов. Однако при рассмотрении такого плана возникают важные этические, статистические и практические вопросы [89, 90].

Независимо от того, являются ли изменения явным образом частью дизайна исследования или возникают в ответ на меняющиеся обстоятельства, важно, чтобы о них сообщалось полностью для помощи читателю в интерпретации результатов. Об изменениях в протоколах в настоящее время не сообщается. Обзор сравнений с протоколами показал, что около половины журнальных статей, описывающих РКИ, содержали необъяснимые расхождения в первичных результатах [57]. Частые необъяснимые расхождения также наблюдались относительно рандомизации, ослепления [91] и методов статистического анализа [92].

Пункт 4а. Критерии включения для участников

Пример: «К исследованию допускались лица старше 18 лет с ВИЧ, которые соответствовали критериям включения для антиретровирусной терапии в соответствии с национальными рекомендациями Малави по лечению ВИЧ (клиническая стадия ВОЗ III или IV или любая стадия ВОЗ с количеством CD4 < 250/мкл), а также те, кто начал лечение, с ИМТ < 18,5. Критериями исключения были беременность и лактация или участие в другой программе дополнительного питания» [93].

Объяснение. Подробное описание критериев включения, используемых для отбора участников, необходимо, чтобы помочь читателям интерпретировать исследование. В частности, четкое понимание этих критериев является одним из нескольких элементов, необходимых для принятия решения о том, к кому применимы результаты исследования, то есть обобщаемости (валидности) исследования и его актуальности для клинической практики или практики общественного здравоохранения (см. пункт 21) [94]. Особое значение имеет характеристика метода отбора участников, например, путем обращения по направлению врача или самостоятельное обращение (например, с помощью объявлений). Поскольку критерии включения применяются до рандомизации, они не влияют на внутреннюю валидность исследования, но играют центральную роль в его внешней валидности.

Типичные и общепринятые критерии отбора связаны с характером и стадией изучаемого заболевания, исключением лиц, считающихся особенно подверженными риску вреда от участия в исследовании, а также с вопросами, необходимыми для обеспечения того, чтобы исследование удовлетворяло правовым и этическим нормам. Например, в экспериментальных исследованиях обычно требуется информированное согласие участников исследования. Разделение критериев включения и исключения обычно не является необходимым; один и тот же критерий

можно сформулировать для включения или исключения участников [95].

Несмотря на важность критериев включения, они зачастую не характеризуются должным образом. В восьми опубликованных РКИ, результаты которых вызвали публикацию клинических предупреждений Национальными институтами здравоохранения, в среднем применялся 31 критерий включения, однако в журнальных статьях было упомянуто только 63% из этих критериев, а в клинических предупреждениях — только 19% [96]. Аналогичные недостатки были обнаружены и в клинических исследованиях ВИЧ [97]. Также, например, критерии включения не были указаны в 25% из 364 публикаций результатов РКИ в хирургии [98].

Пункт 4б. Условия и местоположение, где производился сбор данных

Пример: «Исследование проводилось в клинике антиретровирусной терапии Центральной больницы королевы Елизаветы в Блантайре, Малави, с января 2006 года по апрель 2007 года. Распространенность ВИЧ среди взрослых в 2004 году составляла 27%» [93].

Объяснение. Наряду с критериями включения участников (см. пункт 4а) и описанием вмешательств (см. пункт 5), информация об условиях и месте проведения имеет решающее значение для оценки применимости и обобщаемости исследования. Были ли участники набраны из учреждений первичной, вторичной или специализированной медицинской помощи или из общего населения? Лечебно-профилактические учреждения существенно различаются по принципам организации, квалификации персонала, ресурсам и исходному риску исследуемого заболевания в этой местности? На внешнюю валидность исследования могут также воздействовать климат и другие физические влияния, экономические, географические и социально-культурные факторы окружающей среды.

Авторы должны сообщать о количестве и типе учреждений и медицинских работников, участвующих в проведении испытания. Также должны быть представлены описания учреждений, в которых проводилось исследование, и их расположение с указанием страны, города и непосредственных условий оказания медицинской помощи (например, территориальная медицинская служба, врачебный кабинет, клиническая больница, стационар). В частности, должно быть ясно, проводилось ли исследование в одном или нескольких центрах (многоцентровые исследования). Описание должно содержать достаточно информации, чтобы читатели могли судить о том, могут ли результаты исследования быть применимы в условиях их практики. Среда, в которой проводится исследование, может значительно отличаться от обстановки, в которой результаты исследования впоследствии будут использоваться [94, 99]. Авторы также должны сообщать любую дополнительную информацию об условиях и месте проведения исследования, которые

могли повлиять на результаты наблюдения, например о проблемах с транспортировкой, которые могли повлиять на участие пациентов или задержки в проведении вмешательств.

Пункт 5. Описание вмешательств для каждой группы с приведением деталей, достаточных для воспроизведения, включая то, как и когда они фактически применялись

Примеры: «В исследовании POISE пациенты получали первую дозу исследуемого препарата (например, пероральный Метопролол пролонгированного действия 100 мг или соответствующее плацебо) за 2–4 часа до операции. Для введения исследуемого препарата требовалась частота сердечных сокращений 50 ударов в минуту или более и уровень систолического артериального давления 100 мм рт. ст. или более; гемодинамика проверялась перед каждым введением. Если в любое время в течение первых 6 часов после операции частота сердечных сокращений составляла 80 ударов в минуту или более, а систолическое артериальное давление составляло 100 мм рт. ст. или выше, пациенты получали первую послеоперационную дозу (Метопролол пролонгированного действия 100 мг или соответствующее плацебо) перорально. Если исследуемый препарат не вводили в течение первых 6 часов, пациенты получали первую послеоперационную дозу через 6 часов после операции. Через 12 часов после первой послеоперационной дозы пациенты начинали принимать Метопролол пролонгированного действия перорально по 200 мг или плацебо каждый день в течение 30 дней. Если частота сердечных сокращений пациента постоянно была ниже 45 ударов в минуту или его систолическое артериальное давление падало ниже 100 мм рт. ст., прием исследуемого препарата отменяли до восстановления частоты сердечных сокращений или уровня систолического артериального давления; затем возобновляли прием исследуемого препарата в дозе 100 мг один раз в день. Пациенты, у которых частота сердечных сокращений постоянно составляла 45–49 ударов в минуту, а систолическое артериальное давление превышало 100 мм рт. ст., откладывали прием исследуемого препарата на 12 часов» [100].

«Пациенты были случайным образом распределены для получения изготовленной на заказ неопреновой шины для ношения на ночь или для обычного ухода. Шина представляла собой жесткий ортез для отдыха, рекомендованный для использования только в ночное время. Он покрывал основание большого пальца и возвышение тенара, но не запястье (рис. 1). Шины были изготовлены тремя обученными эрготерапевтами, которые подгоняли шину для каждого пациента так, чтобы можно было открыть первую перемычку и разместить большой палец напротив первого длинного пальца. Пациентам рекомендовалось обращаться к эрготерапевту, если они чувствовали, что шина нуждается в регулировке, усиливалась боль при ношении шины или были побочные эффекты (например, эрозия

кожи). Поскольку в этой ситуации никакое лечение не может считаться золотым стандартом, пациенты в контрольной группе и группе вмешательства получали обычную помощь по усмотрению своего врача (врача общей практики или ревматолога). Мы решили не использовать плацебо, потому что, насколько нам известно, ни одно плацебо для шинирования не обеспечивает успешного ослепления пациентов, как это рекомендуется» [101].

Объяснение. Авторы должны подробно описывать каждое вмешательство, в том числе вмешательства, применявшиеся в группе контроля. Описание должно позволить врачу-клиницисту, желающему использовать вмешательство, осуществленное в исследовании, иметь точные знания в его проведении [102]. Информация о медикаментозном вмешательстве должна включать название препарата, дозу, способ введения (например, перорально, внутривенно), время и продолжительность введения, условия, при которых вмешательство не проводится, и режим титрования, если применимо. Особенно важно детально описывать проводившееся в группе контроля «стандартное лечение» и те вмешательства, которые на самом деле являются комбинацией вмешательств. Авторы должны предоставить подробное описание каждого вмешательства, объяснение порядка, в котором комбинация вмешательств вводится или отменяется, и показаний для их введения, если применимо.

Конкретные расширения заявления CONSORT касаются статей о немедикаментозных и травяных вмешательствах и их конкретных требований к отчетности (таких как опыт, подробности того, как вмешательства были стандартизированы) [43, 44]. Мы рекомендуем читателям обращаться к заявлениям о нефармакологических вмешательствах по мере необходимости.

Пункт 6а. Полностью определенные, предварительно конкретизированные первичные и вторичные показатели выводов, включая указание на то, как и когда они были оценены

Пример: «Первичной конечной точкой в отношении эффективности лечения при псориазе была доля пациентов, у которых за 12 недель было достигнуто уменьшение активности псориаза на 75% от исходного уровня при оценке по индексу PASI [psoriasis area and severity index — индекс распространенности и тяжести псориаза]. Были проведены дополнительные анализы процентного изменения показателей PASI и улучшения состояния псориазических поражений кожи, на которые проводилось воздействие» [103].

Объяснение. Во всех РКИ оцениваются «переменные ответа» или исходы (конечные точки), по которым сравниваются группы. Большинство исследований имеют несколько результатов, некоторые из которых представляют больший интерес, чем другие. Мерой оценки первичного результата является предварительно определенный клинический исход, который имеет наибольшее значение для соответствующих

заинтересованных сторон (таких как пациенты, руководители, врачи, спонсоры) и, как правило, используется для расчета размера выборки (см. пункт 7). Некоторые исследования могут иметь более одного основного исхода. Однако наличие нескольких первичных исходов влечет за собой проблемы интерпретации, связанные с проведением множественных анализов (см. пункты 18 и 20), и поэтому не рекомендуется. Первичные исходы должны быть четко перечислены в статье о РКИ. Другие результаты, представляющие интерес, являются вторичными клиническими исходами (дополнительные исходы). Может быть несколько вторичных результатов, которые часто включают в себя непредвиденные или непредусмотренные последствия изучаемых вмешательств (см. пункт 19), хотя вред всегда следует рассматривать как значимый результат, независимо от того, обозначен ли он как первичный или вторичный.

Все показатели результатов, как первичных, так и вторичных, должны быть перечислены и полностью определены. Принцип здесь заключается в том, что предоставленной информации должно быть достаточно, чтобы позволить другим использовать те же результаты [102]. Когда частоту развития клинического исхода оценивают неоднократно в нескольких моментах времени после рандомизации, авторы должны указать предварительно определенный момент времени, представляющий первостепенный интерес. Для многих немедикаментозных вмешательств полезно указать, кто оценивал клинические исходы (например, если для этого требуются специальные навыки) и сколько специалистов проводили оценку результатов [43].

Там, где это возможно и уместно, следует сообщать об использовании ранее разработанных и утвержденных шкал или согласованных руководств как для повышения качества измерения [104, 105], так и для облегчения сравнения с аналогичными исследованиями [106]. Например, использование способов оценки качества жизни с подтвержденной достоверностью, скорее всего, может улучшить представление результатов соответствующих исследований [107]. Авторы должны указать источники, в которых описана шкала оценки, и кратко охарактеризовать принципы ее применения.

В 196 РКИ, рассматривавших эффективность нестероидных противовоспалительных препаратов при ревматоидном артрите, было оценено более 70 исходов [108], а в 2000 исследованиях, проведенных у больных шизофренией, использовалось 640 различных инструментов оценки, из которых 369 применялись впервые [33]. Исследование 149 из этих 2000 испытаний показало, что неопубликованные ранее шкалы были источником систематической ошибки. В исследованиях нефармакологических методов лечения треть заключений о превосходстве того или иного метода была сделана на основании оценки с использованием ранее не опубликованных шкал и не появилась бы, если бы авторы применяли опубликованные проверенные шкалы оценки [109]. О подобных

фактах сообщается и в других публикациях [110, 111]. Только в 45% из 519 РКИ, опубликованных в 2000 году, указан первичный исход [16]; это сопоставимо с 53% для аналогичной когорты из 614 РКИ, опубликованных в 2006 г. [17].

Пункт 6б. Любые изменения результатов исследования после его начала с указанием причин

Пример: «Первоначальной первичной конечной точкой была смертность от всех причин, но во время скрытого анализа комиссия по мониторингу данных и безопасности отметила, что общая смертность была ниже, чем предполагалось, и что исследование не может быть завершено с размером выборки и мощностью, которые изначально планировались. Поэтому руководящий комитет решил ввести дополнительные первичные конечные точки: смертность от всех причин (исходная первичная конечная точка) вместе со смертностью от всех причин или госпитализацией по сердечно-сосудистым заболеваниям (первая предварительно определенная вторичная конечная точка)» [112].

Объяснение. Существует множество причин отклонения от первоначального протокола исследования (см. пункт 24). Авторы должны сообщать обо всех основных изменениях в протоколе, включая незапланированные изменения критериев приемлемости, вмешательств, обследований, сбора данных, методов анализа и исходов. Такая информация не всегда сообщается.

Как указывалось ранее (см. пункт 6а), в большинстве исследований регистрируется несколько исходов с риском того, что результаты будут представлены только для выбранной подгруппы (см. пункт 17). Предварительная спецификация и отчетность о первичных и вторичных исходах (см. пункт 6а) должны устранить такой риск. Однако в некоторых исследованиях обстоятельства требуют изменения способа оценки исхода или даже, как в приведенном выше примере, перехода к другому исходу. Например, могут быть внешние данные из других исследований или систематических обзоров, предполагающие, что конечная точка может быть неадекватной, или набор участников или общая частота событий в исследовании могут быть ниже, чем ожидалось [112]. Изменение конечной точки в случае исследований без ослепления данных гораздо более проблематично, хотя это может быть указано в контексте адаптивного дизайна исследования [88]. Авторы должны определить и объяснить любые такие изменения. Аналогичным образом следует сообщать и объяснять любые изменения после начала исследования в отношении определения исходов — первичных или вторичных.

Сравнение протоколов и публикаций 102 рандомизированных исследований показало, что в 62% статей об исследованиях по крайней мере один первичный исход был изменен, введен или опущен по сравнению с протоколом [55]. Первичные исходы также различались между протоколами и публикациями для 40%

из 48 испытаний, финансируемых Канадским институтом медицинских исследований [113]. Ни в одной из последующих 150 статей об исследованиях не упоминались, не говоря уже о пояснениях, изменения протокола. Аналогичные результаты других исследований были недавно опубликованы в систематическом обзоре эмпирических исследований, посвященных систематической ошибке в отчетах о результатах [57].

Пункт 7а. Определение размера выборки

Примеры: «Для выявления снижения PHS (postoperative hospital stay — послеоперационное пребывание в стационаре) на 3 дня (SD 5 дней), что согласуется с исследованием Lobo et al [17], при двустороннем уровне значимости 5% и мощностью 80%, был необходим размер выборки 50 пациентов на группу, учитывая ожидаемый уровень отсева 10%. Для набора такого количества пациентов предполагался 12-месячный период включения» [114].

«Основываясь на ожидаемой частоте первичной комбинированной конечной точки в 11% через 2,25 года в группе плацебо, мы подсчитали, что нам потребуется 950 событий первичной конечной точки и размер выборки 9650 пациентов, чтобы обеспечить 90% мощности для выявления значимой разницы между Ивабрадином и плацебо, соответствующей 19% снижения относительного риска (с двусторонней ошибкой 1-го типа 5%). Первоначально мы разработали исследование, управляемое событиями, и планировали прекратить его, когда произойдет 950 событий первичной конечной точки. Однако частота первичной конечной точки была выше, чем прогнозировалось, возможно, из-за исходных характеристик включенных пациентов, которые имели более высокий риск, чем ожидалось (например, более низкая доля пациентов с классом I по NYHA и более высокие показатели диабета и гипертонии). Мы подсчитали, что при возникновении 950 первичных событий конечной точки последние включенные пациенты получали лечение только около 3 месяцев. Поэтому в январе 2007 года исполнительный комитет решил изменить исследование с событийного на временное и продолжить исследование до тех пор, пока пациенты, которые были рандомизированы последними, не будут наблюдаться в течение 12 месяцев. Это изменение не повлияло на запланированную продолжительность исследования в 3 года» [115].

Объяснение. Исходя из научных и этических принципов, размер выборки для исследования необходимо тщательно планировать, соблюдая баланс между клиническими и статистическими соображениями. В идеале исследование должно быть достаточно большим, чтобы с высокой вероятностью (статистической мощностью) выявить как статистически, так и клинически значимое различие заданного размера, если такое различие существует. Размер эффекта, который считается клинически важным, обратно пропорционален размеру выборки, необходимой для его выявления,

то есть для обнаружения небольших различий необходимы большие выборки. Элементами расчета размера выборки являются: 1) предполагаемые исходы в каждой группе (при предполагаемом клинически значимом различии между группами сравнения); 2) уровень ошибки α (ошибки типа I); 3) статистическая мощность исследования (или уровень ошибки β (ошибки типа II); 4) стандартное отклонение измерений при регистрации непрерывных результатов [116]. Взаимодействие этих элементов и их отчеты будут отличаться для кластерных исследований [40] и исследований не меньшей эффективности и эквивалентности [39].

Авторы должны указать, как был определен размер выборки. Если использовался общепринятый расчет статистической мощности, авторы должны указать первичный результат, на котором был основан расчет (см. пункт 6а), все количественные показатели, использованные в расчете, и расчетный размер выборки для каждой группы сравнения. Желательно представить ожидаемый результат в контрольной группе и разницу между группами, которую не хотелось бы упускать из виду. В качестве альтернативы авторы могут представить процентное соотношение события или среднего значения для каждой группы, используемой в их расчетах. Должна быть подробно описана любая поправка на отсев или несоблюдение требований во время исследования.

Некоторые методологи писали, что так называемые исследования с недостаточной мощностью могут быть приемлемыми, потому что в конечном итоге их можно объединить в систематический обзор и метаанализ [117–119], а также потому, что любая информация лучше, чем ее отсутствие. Следует отметить важные оговорки, например, исследование должно быть беспристрастным, должным образом публиковаться независимо от результатов, чтобы оно стало доступным для метаанализа [118]. С другой стороны, многие исследователи-медики опасаются, что исследования с недостаточной мощностью и неопределенными результатами останутся неопубликованными, и настаивают на том, что все исследования должны иметь «достаточную мощность». Эти дебаты будут продолжаться, и у членов группы CONSORT разные мнения на этот счет. Какой бы ни была сила исследования, авторы должны правильно указать предполагаемый размер со всеми своими методами и предположениями [118]. Это доступно раскрывает силу исследования для читателей и дает им способ оценить, достигло ли исследование запланированного масштаба.

В некоторых исследованиях для помощи в принятии решения о целесообразности продления набора участников используются промежуточные анализы (см. пункт 7b). Если фактический размер выборки отличался от изначально предполагаемого размера выборки по какой-либо другой причине (например, из-за трудностей набора участников или пересмотра

целевого размера выборки), то должно быть дано объяснение.

Статьи об исследованиях с небольшими выборками часто содержат ошибочный вывод об отсутствии различий между группами, хотя на самом деле было изучено слишком мало пациентов, чтобы делать такие заявления [120]. Обзоры литературы однозначно отмечают, что значительная доля РКИ имеет слишком низкую статистическую мощность для выявления клинически значимых эффектов лечения [121–123]. В действительности возможны небольшие, но клинически ценные истинные различия между группами вмешательства, которые могли бы быть статистически подтверждены при более крупных испытаниях [124].

В целом сообщаемые размеры выборки в исследованиях кажутся небольшими. Средний размер выборки составил 54 пациента в 196 исследованиях при артритах [108], 46 пациентов в 73 исследованиях при кожных заболеваниях [8] и 65 пациентов в 2000 исследованиях при шизофрении [33]. Эти небольшие размеры выборки согласуются с результатами 519 исследований, проиндексированных в PubMed в декабре 2000 года [16], и аналогичной когорты исследований ($n = 616$), проиндексированных в PubMed в 2006 году [17], где среднее число пациентов, набранных для исследований в параллельных группах, составляло 80 за оба года. Более того, во многих обзорах было обнаружено, что лишь немногие авторы сообщают о том, как они определяли размер выборки [8, 14, 32, 33, 123].

Мало смысла в расчете статистической мощности после получения результатов исследования, так как ее достаточно точно отражают доверительные интервалы (см. пункт 17) [125].

Пункт 76. При необходимости объяснение любого промежуточного анализа и оснований для прекращения исследования

Примеры: «В ходе исследования было проведено два промежуточных анализа. Уровни статистической значимости в целом соответствовали значению $P = 0,05$ и были рассчитаны в соответствии с правилом прекращения исследований (правило O'Brien-Fleming). При окончательном анализе использовали Z-критерий, равный 1,985, с соответствующим значением p , равным 0,0471» [126].

«Независимый совет по мониторингу данных и безопасности периодически пересматривал данные об эффективности и безопасности. Правила остановки были основаны на модифицированных границах Хейбитла-Петто: 4 SD в первой половине исследования и 3 SD во второй половине для данных об эффективности, 3 SD в первой половине исследования и 2 SD во второй половине для данных о безопасности. Два формальных промежуточных анализа эффективности были выполнены, когда накопилось 50% и 75% ожидаемого числа первичных событий; никакая коррекция заявленного значения P для этих промежуточных тестов не проводилась» [127].

Объяснение. Для многих исследований набор участников происходит в течение длительного периода времени. Если в ходе исследования выявляются особенно хорошие или, напротив, очень плохие результаты при применении какого-либо вмешательства, может возникнуть необходимость досрочного прекращения исследования по этическим соображениям. Такое решение может быть принято на основании промежуточного анализа по мере накопления данных, предпочтительно независимым комитетом по мониторингу данных [128]. Однако если данные исследования по мере их накопления анализируются в пяти промежуточных анализах, то общая частота ложноположительных результатов составляет почти 19%, а при единственном анализе всех данных — около 5%.

Доступны несколько групповых последовательных статистических методов для корректировки множественных анализов [129–131], и их использование должно быть предварительно указано в протоколе исследования. В ходе каждого промежуточного анализа с помощью этих методов проводится сравнение данных в группах, при этом очень низкие значения P свидетельствуют о статистической значимости различий. Некоторые исследователи используют групповые последовательные методы как помощь в принятии решений [132], тогда как другие рассматривают их как формальное правило остановки (с намерением прекратить исследование, если наблюдаемое значение P меньше критического значения).

Авторы должны сообщить о промежуточных анализах данных, их количестве, применявшихся статистических методах (включая любые правила прекращения исследования) и времени, когда они были запланированы (до начала исследования, до того, как комитет по мониторингу данных увидит какие-либо промежуточные данные по распределению, или позднее). Эта информация часто не включается в опубликованные статьи об исследованиях [133], даже в тех исследованиях, которые сообщают о прекращении раньше, чем планировалось [134].

Рандомизация

Пункт 8а. Метод, используемый для генерации последовательности случайного распределения участников исследования

Примеры: «Независимые фармацевты отпускали либо ингаляторы с активным веществом, либо плацебо в соответствии с составленным компьютером списком рандомизации» [63].

«Для распределения участников использовался сгенерированный компьютером список случайных чисел» [135].

Объяснение: Участники должны быть распределены в группы сравнения в исследовании на основе рандомизации (случайного процесса, характеризующегося непредсказуемостью) (вставка 1). Авторы должны предоставить достаточную информацию, чтобы читатель мог оценить методы, использованные для создания последовательности случайного распределения

в группы сравнения, и вероятность возникновения систематической ошибки на этом этапе. Важно, чтобы информация о процессе рандомизации была включена в основную часть статьи, а не в виде отдельного дополнительного файла, что может быть пропущено читателем.

Термин «рандомизированный» имеет точное математическое значение. При случайном распределении каждый участник имеет известную вероятность получения каждого вмешательства до того, как оно будет применено, но фактически назначенное вмешательство определяется случайным образом и не может быть предсказано. Однако термин «рандомизированный» часто необоснованно используется в литературе для описания исследований, в которых применялись неслучайные, детерминированные методы распределения участников на основании очередности, номеров историй болезни или даты рождения. Когда исследователи используют такие неслучайные методы, они должны точно их описывать и не применять термин «рандомизированный» или любой его синоним. Даже термин «псевдорандомизированный» неприемлем для описания таких исследований. Исследования, основанные на неслучайных методах, обычно дают смещенные (предвзятые) результаты [2–4, 136]. Систематические ошибки, предположительно, возникают из-за невозможности сохранить в тайне принцип распределения участников в группы сравнения (см. пункт 9).

Существует много эффективных методов генерации случайной последовательности. Однако читатели не могут судить об адекватности использованных в РКИ методов только по таким терминам, как «случайное распределение», «рандомизация» или «случайное», без дальнейшего уточнения. Авторы должны указать использованный метод генерации случайной последовательности, например таблица случайных чисел или компьютеризированный генератор случайных чисел. Последовательность может быть получена методом минимизации, неслучайным, но общепринятым методом (вставка 2).

В некоторых исследованиях участников намеренно распределяют в группы в неравных количествах по каждому вмешательству, например, чтобы получить больше опыта применения новой процедуры или ограничить расходы на проведение исследования. В таких случаях авторы должны указать коэффициент рандомизации (например, 2:1 или два участника лечения на каждого участника контроля) (см. пункт 3а).

В репрезентативной выборке исследований, проиндексированных PubMed в 2000 году, только 21% сообщили об адекватном подходе к генерации случайных последовательностей [16]; этот показатель увеличился до 34% для аналогичной когорты индексированных PubMed исследований в 2006 году [17]. Более чем в 90% этих случаев исследователи использовали генератор случайных чисел на компьютере или таблицу случайных чисел.

Вставка 2. Рандомизация и минимизация

•Простая рандомизация — чистая рандомизация, основанная на одном коэффициенте распределения. Простая рандомизация с соотношением распределения 1:1 аналогична подбрасыванию монеты, хотя мы не поддерживаем подбрасывание монеты для рандомизации в РКИ. «Простой» — это несколько неправильное название. В то время как другие схемы рандомизации кажутся сложными и изощренными, на самом деле простая рандомизация элегантно сложна, поскольку она более непредсказуема и превосходит все другие альтернативы по уровню предотвращения смещения.

•Ограниченная рандомизация — любой рандомизированный подход, отличный от простой рандомизации. Заблокированная рандомизация является наиболее распространенной формой. Другие средства ограниченной рандомизации включают замену, смещенную монету и рандомизацию урн, хотя они используются гораздо реже [141].

•Заблокированная рандомизация. Рандомизация внутри блоков используется для обеспечения того, чтобы группы сравнения были созданы в соответствии с заданным соотношением, обычно 1:1, или группами примерно одинакового размера. Рандомизация внутри блоков может быть использована, чтобы обеспечить баланс размеров групп в любом периоде исследования. Например, в каждом блоке из восьми участников по четыре будут распределены в каждую группу исследования [142]. Улучшение баланса достигается за счет уменьшения непредсказуемости последовательности. Хотя вмешательства назначаются в случайном порядке в пределах каждого блока, человек, проводящий исследование, может сделать вывод о некоторых следующих распределениях лечения, если известен размер блока [143]. Ослепление вмешательств, использование больших размеров блоков и случайное изменение размера блоков может решить эту проблему.

•Стратифицированная рандомизация используется для обеспечения хорошего баланса характеристик участников в каждой группе. При случайном распределении, особенно в небольших исследованиях, группы сравнения могут оказаться неоднородными по некоторым исходным характеристикам участников, например по возрасту и стадии заболевания. Это снижает надежность результатов исследования [144]. Такого дисбаланса можно избежать, не жертвуя преимуществами рандомизации. Стратификация гарантирует, что количество участников, получающих каждое вмешательство, точно сбалансировано внутри каждой страты (подгруппы). Стратифицированная рандомизация достигается путем проведения отдельной процедуры рандомизации в каждой из двух или более подгрупп участников, выделенных

по определенному признаку (например, среди лиц, которые определяют каждый исследовательский центр, лиц определенного возраста или с определенной тяжестью заболевания). Стратификация по исследовательским центрам распространена в многоцентровых исследованиях. Стратификация требует некоторой формы ограничения (например, блокировки внутри страт). Стратификация без блокировки малоэффективна.

•Минимизация обеспечивает баланс между группами вмешательства по нескольким характеристикам пациентов (таким как возраст) [22, 60]. Только первый пациент действительно распределяется случайным образом, а для каждого последующего участника лечение определяют так, чтобы свести к минимуму имеющийся в этот момент дисбаланс по выбранным факторам между группами. В дальнейшем может быть использован этот подход к распределению в группы лечения или может быть сделан случайный выбор в пользу вмешательства, которое может уменьшить неоднородность групп (например, с вероятностью, равной 0,8). Обычно предпочтительнее использовать случайную составляющую. Преимущество минимизации состоит в том, что малые группы получают очень похожими по характеристикам участников на всех этапах исследования.

Минимизация предлагает единственную приемлемую альтернативу рандомизации, и некоторые утверждают, что этот метод эффективнее рандомизации [145]. С другой стороны, минимизация не имеет теоретической основы для устранения предвзятости по всем известным и неизвестным факторам. Тем не менее в целом исследования, в которых используется метод минимизации, считаются методологически эквивалентными рандомизированным исследованиям, даже без применения элемента случайности.

Пункт 86. Тип рандомизации; детализация любых ограничений (таких как объединение в группы и размер групп)

Примеры: «Последовательность рандомизации была создана с использованием статистического программного обеспечения Stata 9.0 (StataCorp, College Station, TX) и стратифицирована по центрам с распределением 1:1 с использованием случайных размеров блоков 2, 4 и 6» [137].

«Участники были случайным образом распределены в соответствии с простыми процедурами рандомизации (компьютеризированные случайные числа) в 1 из 2 групп лечения» [138].

Объяснение. В исследованиях с участием нескольких сотен или более участников обычно можно доверять простой рандомизации для формирования одинаковых чисел в двух группах исследований [139] и для создания групп, которые примерно сопоставимы с точки зрения известных и неизвестных

прогностических переменных [140]. Для небольших исследований (см. пункт 7а) и даже для исследований, которые не должны быть небольшими, поскольку они могут прекратиться до достижения необходимого размера, может быть полезна некоторая ограниченная рандомизация чтобы сбалансировать группы по размеру или характеристикам (вставка 2).

Важно отметить, не использовались ли какие-либо дополнительные методы ограничения, указав их или заявив, что была проведена «простая рандомизация». В противном случае следует указать методы, используемые для ограничения рандомизации, а также метод, используемый для генерации случайной последовательности. При использовании блоковой рандомизации авторы должны предоставить подробную информацию о том, как блоки были сгенерированы (например, с использованием конструкции переставленных блоков с компьютерным генератором случайных чисел), какой был размер блоков, а также информацию о том, был ли размер блоков фиксированным или варьировался случайным образом. Если исследователи узнают о размере(ах) блока, необходимо сообщить об этом, так как данная информация может привести к дешифрованию. При проведении стратифицированной рандомизации авторам следует перечислить принципы составления подгрупп и блоков (такие как место выборки, пол, стадия заболевания), пороговые значения категоризации в пластах и метод, использованный для ограничения. Хотя стратификация является эффективным методом, особенно для небольших исследований, однако ее сложно реализовать и она может оказаться невозможной при учете большого числа признаков. Авторы должны четко сообщать о применении процесса минимизации (вставка 2), переменных, включенных в схему, и использовании элемента случайности.

Только в 9% из 206 статей об исследованиях в специализированных журналах [23] и в 39% из 80 статей об исследованиях в общих медицинских журналах сообщали об использовании стратификации [32]. В каждом случае только около половины статей упоминали об использовании ограниченной рандомизации. Однако эти исследования, а также исследования, по данным Adetugbo и Williams [8], показали, что во многих статьях о РКИ, авторы которых не упоминают о применении блокирования или стратификации, размеры групп часто одинаковые или почти одинаковые. Одним из возможных объяснений близкого баланса в числах является занижение сведений об использовании ограниченной рандомизации.

Пункт 9. Механизм, используемый для осуществления случайной последовательности распределения участников в группы исследования (такой как последовательно пронумерованные контейнеры), описывающий любые шаги,

предпринятые для сокрытия этой последовательности до тех пор, пока не будут назначены вмешательства

Примеры: «Доксициклин и плацебо были в форме капсул и идентичны по внешнему виду. Они были предварительно упакованы во флаконах и последовательно пронумерованы для каждой женщины в соответствии с графиком рандомизации. Каждой женщине был присвоен порядковый номер, и они получали капсулы в соответствующем расфасованном флаконе» [146].

«Последовательность распределения была скрыта от исследователя (JR), который регистрировал и оценивал участников, с помощью последовательно пронумерованных, непрозрачных, запечатанных и скрепленных конвертах. Внутри конверта использовалась алюминиевая фольга, чтобы сделать конверт непроницаемым для интенсивного света. Чтобы предотвратить нарушение порядка распределения, имя и дата рождения участника были написаны на конверте, а из запечатанного конверта была сделана видеокассета с видимыми данными участника. Копировальная бумага внутри конверта перенесла информацию на карточку распределения внутри конверта, а второй исследователь (CC) позже просмотрел видеокассеты, чтобы убедиться, что конверты все еще были запечатаны, когда на них были написаны имена участников. Соответствующие конверты открывались только после того, как зачисленные участники завершили все базовые оценки и пришло время распределять вмешательство» [147].

Объяснение. В пункте 8а обсуждается процесс генерации непредсказуемой последовательности распределения участников в группы исследования. Большое значение имеет то, как эта последовательность применяется при включении участников в исследование (вставка 1). Сгенерированный график распределения должен быть реализован с использованием сокрытия распределения [23] — критически важного механизма, который позволяет сохранять назначенное лечение в тайне от исполнителей и, таким образом, предупреждает возможное субъективное воздействие на результаты исследования. При принятии решения о включении или отклонении от включения участника исследования, а также при получении информированного согласия участника нет необходимости учитывать, в какую группу будет отнесен следующий участник [148].

Сокрытие распределения не следует путать с ослеплением (см. пункт 11). Сокрытие распределения направлено на предотвращение систематической ошибки, связанной с отбором участников, сохраняет в тайне последовательность их отнесения к той или иной группе до момента распределения и всегда может быть успешно реализовано [2]. Напротив, ослепление предназначено для предотвращения систематических ошибок, связанных с ходом исследования или выявлением и подтверждением определенного клинического исхода после распределения в группы, и далеко не всегда может быть реализовано [23].

Однако без надлежащего сокрытия распределения даже случайные, непредсказуемые последовательности распределения могут быть нарушены [2, 149].

Особенно желательно применение централизованного распределения или распределения «третьим лицом». Многие эффективные подходы сокрытия распределения включают привлечение независимого внешнего вмешательства. Использование техники централизованной рандомизации по телефону или через аптеку — два распространенных метода. Автоматизированные системы назначения, вероятно, станут более распространенными [150]. Когда рандомизированное распределение с внешним участием невозможно, отличным методом сокрытия распределения является использование пронумерованных контейнеров. Вмешательства (часто непосредственно лекарства) запечатаны в последовательно пронумерованные идентичные контейнеры в соответствии с последовательностью распределения [151]. Вложение назначений в последовательно пронумерованные, непрозрачные, запечатанные конверты может быть хорошим механизмом сокрытия распределения, если он тщательно разрабатывается и контролируется. Однако этот метод может оказаться неэффективным при плохом качестве выполнения. Исследователи должны убедиться, что конверты непрозрачны, когда их держат на свету, а вскрытие конвертов производится последовательно и только после того, как на них были написаны имя участника и другие необходимые данные [143].

Ряд методологических исследований предоставляет эмпирические данные в поддержку этих мер предосторожности [152, 153]. Исследования, в которых последовательность распределения была неадекватно или нечетко скрыта, дали более высокие оценки эффектов лечения, чем исследования, в которых авторы сообщали об адекватном сокрытии распределения. Эти результаты обеспечивают убедительные эмпирические доказательства того, что неадекватное сокрытие распределения способствует систематической ошибке в оценке эффектов лечения.

Несмотря на важность механизма сокрытия распределения, в опубликованных статьях такие подробности часто не упоминаются. Механизм, используемый для распределения вмешательств, был описан в 89% статей об исследованиях по ревматоидному артриту [108], в 48% статей о РКИ в журналах по акушерству и гинекологии [23] и в 44% статей об исследованиях в общих медицинских журналах [32]. В более широкой репрезентативной выборке из всех рандомизированных исследований, проиндексированных в PubMed, только 18% сообщили о каком-либо механизме сокрытия распределения, но некоторые из этих описанных механизмов были неадекватными [16].

Пункт 10. Разработка случайной последовательности распределения участников в группы исследования, регистрация участников, распределение участников по группам вмешательств

Примеры: «Определение того, каким больным будут назначены терапия стрептомицином и постельный

режим (группа S) или только постельный режим (контрольная группа C), проводилось с помощью порядка рандомизации, основанного на результатах процесса генерации случайных чисел, который профессор Бредфорд Хилл проводил отдельно для мужчин и женщин в каждом медицинском центре. Информация о рандомизации была неизвестна ни исследователям, ни координатору исследования... После одобрения специальной комиссией включения пациента, но еще до поступления пациента в центр лечения стрептомицином в центральном офисе вскрывали конверт с соответствующим номером; на карте внутри конверта было указано, в какую группу лечения отнесен больной (группу S или группу C), затем эту информацию передавали медицинскому работнику центра» [24].

«Сведения о выделенной группе были даны на цветных карточках, помещенных в последовательно пронумерованные, непрозрачные, запечатанные конверты. Они были подготовлены в NPEU (National Perinatal Epidemiology Unit — Национальное подразделение перинатальной эпидемиологии) и хранились в согласованном месте в каждой палате. Рандомизация проводилась в конце 2-го периода родов, когда акушерка информировала о начале естественных родов. Для включения женщины в исследование акушерка открывала следующий последовательно пронумерованный конверт» [154].

«Заблокированная рандомизация проводилась с помощью составленного компьютером списка случайных чисел, подготовленного исследователем, не участвовавшим в клиническом исследовании. Мы распределяли на группы пациентов, поступающих для проведения процедуры, связанной с онкологическим заболеванием. После того как медсестра-исследователь получала согласие пациента, она связывалась с контактным лицом, которое не было связано с процессом распределения участников исследования на группы» [155].

Объяснение. Как отмечено в пункте 9, сокрытие назначенного вмешательства на этапе включения участников в исследование особенно важно. Таким образом, помимо методов сокрытия распределения важно знать, как осуществлялась рандомизация пациентов, в частности, кто создавал случайную последовательность распределения, проводил набор участников и их распределение по группам вмешательств.

Процесс рандомизации участников исследования состоит из трех различных этапов: создание случайной последовательности, сокрытие распределения и реализация процесса (вставка 3). Хотя в каждом из этих процессов один исследователь может выполнять несколько задач, следует стремиться к полному разделению людей, занимающихся генерацией и сокрытием распределения, от людей, занимающихся выполнением заданий. Таким образом, если кто-то участвует в этапах генерации последовательности или сокрытия распределения, в идеале он не должен участвовать в этапе реализации.

Даже при безупречном качестве процесса рандомизации неспособность разделить создание и сокрытие последовательности распределения от собственно

процесса реализации может привести к систематической ошибке. Например, исследователь, участвовавший в процессе генерации последовательности распределения участников в группы вмешательств, может сохранить копию и воспользоваться ею при беседе с потенциальными участниками исследования. Таким образом, это может повлиять на процесс включения пациентов в исследование или назначения им лечения, несмотря на то, что последовательность была непредсказуемой. Также исследователи должны не только обеспечить непредсказуемость последовательности распределения, но и хранить ее недоступно (например, в сейфе в здании, довольно недоступном для места регистрации) даже для лиц, которые ее создали. В статье об исследовании должно быть указано, где исследователи хранили список распределения.

Вставка 3. Этапы типичного процесса рандомизации

Генерация случайной последовательности

- Сгенерируйте последовательность распределения некоторой случайной процедурой

Соккрытие распределения

- Разработать механизм сокрытия распределения (например, пронумерованные идентичные емкости или последовательно пронумерованные запечатанные непрозрачные конверты)
- Подготовить механизм сокрытия распределения, используя последовательность распределения, полученную на этапе генерации последовательности

Создание случайной последовательности

- Включение участников:
 - оценка соответствия кандидата критериям включения;
 - обсуждение исследования с пациентом;
 - получение информированного согласия больного;
 - включение больного в исследование
- Раскрытие информации об отнесении участника к той или иной группе (например, вскрытие контейнера или конверта)
- Применение назначенного вмешательства

Пункт 11а. Если было проведено слепое исследование, указать, кто был ослеплен после определения в группы вмешательств (например, участники, исполнители, лица, оценивающие исходы) и каким образом

Примеры: «В то время как пациенты и врачи, отнесенные к группе вмешательства, знали о выделенной группе, лица, оценивающие результаты исследования, и аналитики данных не знали о распределении» [156].

«Ослепление и уравниловность строго поддерживались, подчеркивая персоналу и участникам

вмешательства, что каждая диета придерживается принципов здорового образа жизни и каждая из них, по мнению некоторых экспертов, лучше подходит для долгосрочной потери веса. За исключением интервенционистов (диетологов и психологов-бихевиористов), исследователи и сотрудники не знали о назначении участникам диеты. В исследовании придерживались установленных процессов для сохранения разделения между исследователями, которые оценивают результаты, и исследователями, которые осуществляют вмешательство. Сотрудники, получившие результаты измерений, не были проинформированы о распределении диеты по группам. Исследователи, проводившие вмешательство, диетологи и поведенческие психологи не оценивали результаты. Все исследователи, персонал и участники исследования были замаскированы для оценки результатов исследования» [157].

Объяснение: Термин «ослепление» или «маскирование вмешательств» означает сокрытие информации о назначенных вмешательствах от людей, участвующих в исследовании, на которых эта информация потенциально может повлиять. Ослепление — важная защита от предвзятости, особенно при оценке субъективных результатов [153].

Бенджамин Франклин считается первым, кто использовал ослепление в научном эксперименте [158]. Он завязал участникам глаза, чтобы они не знали, когда он применяет месмеризм (популярную «целебную жидкость» XVIII века), и тем самым показал, что месмеризм — это обман. Основываясь на этом эксперименте, научное сообщество признало способность ослепления уменьшать предвзятость, и эта стратегия остается широко используемой стратегией в научных экспериментах.

В вставке 4 в терминологии ослепления определяются группы лиц (такие как участники исследования, исполнители, лица, участвующие в сборе данных, оценивающие результаты исследования, и аналитики данных), которые потенциально могут внести предвзятость в исследование, зная распределение вмешательств. Пациенты, которые знают о своем назначении лечения, могут реагировать по-разному (например, реагировать более благоприятно, если они получают новое лечение) [153]. Отсутствие ослепления также может повлиять на соблюдение режима вмешательства, использование сопутствующих вмешательств и риск выбывания из исследования.

Неослепленные исследователи могут вводить аналогичные предубеждения, а неослепленные лица, участвующие в сборе данных, могут по-разному оценивать результаты (например, частоту или время), повторять измерения аномальных результатов или оказывать поддержку при проведении тестирования. Неослепленные лица, оценивающие результаты исследования, могут по-разному оценивать субъективные результаты, а неослепленные аналитики данных могут внести предвзятость за счет выбора аналитических стратегий, таких как выбор благоприятных моментов времени или результатов, а также путем принятия решений об исключении пациентов из анализа.

Эти предубеждения были хорошо задокументированы [71, 153, 159–162].

Ослепление, в отличие от сокрытия распределения (см. пункт 10), не всегда может быть целесообразным или возможным. Примером может служить исследование, в котором сравнивалась интенсивность боли, связанной с забором крови из мочки уха или пальца [163]. Ослепление особенно важно, когда оценка исхода включает некоторую субъективность, например при оценке боли. Ослепление лиц, регистрирующих данные, и лиц, выносящих решения, вряд ли будет иметь значение для объективных исходов, таких как смерть по любой причине. Однако даже в этом случае отсутствие ослепления участников или исследователей может привести к другим проблемам, таким как отсев данных [164]. В некоторых исследованиях, особенно хирургического профиля, применение двойного слепого метода (ослепление участников исследования и хирургов) часто затруднено или невозможно, но ослепление лиц, осуществляющих сбор данных, и лиц, оценивающих результаты, часто достижимо. Например, фотографии поражения могут быть сделаны до и после лечения и оценены внешним наблюдателем [165]. Независимо от того, возможно ли ослепление, авторы могут и должны всегда указывать, кто был ослеплен (то есть участники, медицинские работники, лица, участвующие в сборе данных или оценивающие результаты исследования).

К сожалению, авторы часто не сообщают, применялось ли ослепление [166]. Так, информации о применении слепого метода нет в 51% из 506 отчетов о РКИ при муковисцидозе [167], в 33% из 196 отчетов о РКИ у больных ревматоидным артритом [108] и в 38% из 68 отчетов о РКИ в дерматологии [8]. До тех пор пока авторы исследований не улучшат данные об использовании ослепления, читателям будет трудно судить о достоверности исследований, которые они могут захотеть использовать в своей клинической практике.

Термин «маскирование» иногда используется вместо «ослепления», чтобы избежать путаницы с медицинским состоянием отсутствия зрения (слепота). Тем не менее термин «ослепление» в методологическом смысле понятен во всем мире и тоже приемлем для представления статей о РКИ [165, 168].

Вставка 4. Терминология, используемая в методе ослепления

Чтобы научный термин имел полезность, он должен быть последовательным в своем использовании и толковании. Авторы исследований обычно используют термин «двойной слепой» и реже термин «простой слепой» или «тройной слепой». Проблема с этим лексиконом заключается в том, что существуют большие различия в интерпретациях этих терминов клиницистами и определениях этих терминов в эпидемиологических учебниках [169]. Более того, исследование 200 РКИ, описанных

как двойные слепые, выявило 18 различных комбинаций групп, фактически ослепленных, когда авторы этих исследований были опрошены, и примерно в одном из каждых пяти этих исследований, зарегистрированных как двойные слепые, не ослепляли ни участников, ни медицинских работников, ни сборщиков данных [170].

Это исследование показывает, что термины неоднозначны, и поэтому авторам и редакторам следует отказаться от их использования. Вместо этого авторы должны явно сообщать об ослеплении людей, для которых ослепление может повлиять на достоверность исследования.

К медицинским работникам относится весь персонал (например, врачи, мануальные терапевты, физиотерапевты, медсестры), который ухаживает за участниками во время исследования. Регистраторы данных — это лица, которые собирают данные о результатах исследований. Ответственные за результаты — это лица, которые определяют, имел ли участник какие-либо последствия, представляющие интерес.

Некоторые исследователи также выступали за ослепление и сообщали об ослеплении комитету по мониторингу данных и авторам статей [160]. Ослепление этих групп встречается редко, и необходимость их ослепления обсуждается [171].

Иногда одна группа лиц (например, представители здравоохранения) — это те же лица, выполняющие другую роль в исследовании (например, регистраторы данных). Даже если это так, авторы должны прямо указать ослепляющий статус этих групп, чтобы читатели могли судить о достоверности исследования.

Пункт 11б. Если применимо, описание сходства вмешательств

Пример: «Jamieson Laboratories Inc. предоставила 500 мг Ниацина с немедленным высвобождением в белых, продолговатых, разделенных пополам таблетках. Мы независимо друг от друга подтвердили содержание таблеток с помощью высокоэффективной жидкостной хроматографии... Плацебо соответствовало исследуемому препарату по вкусу, цвету и размеру и содержало микрокристаллическую целлюлозу, диоксид кремния, дикальцийфосфат, стеарат магния и стеариновую кислоту» [172].

Объяснение. Подобно тому, как мы ищем доказательства в действительно случайном назначении вмешательства, мы стремимся получить доказательства эффективности метода ослепления. В исследованиях с ослеплением участников или медицинских работников авторы должны указать сходство характеристик вмешательств (таких как внешний вид, вкус, запах и способ введения) [35, 173].

Некоторые люди выступают за тестирование на ослепление, спрашивая участников или медицинских работников в конце исследования, считают ли они, что участник получил экспериментальное или

контрольное вмешательство [174]. Поскольку участники и исследователи обычно знают, прошел ли участник первичный результат, это затрудняет определение того, отражают ли их ответы неудачу ослепления или точные предположения об эффективности вмешательства [175]. Учитывая неопределенность, связанную с этим типом информации, мы убрали из заявления CONSORT 2010 рекомендацию сообщать об этом типе тестирования для ослепления. Однако мы выступаем за то, чтобы авторы сообщали обо всех известных компромиссах в ослеплении. Например, авторы должны сообщать, было ли необходимо освободить кого-либо из участников в любой момент проведения исследования.

Пункт 12а. Статистические методы, используемые для сравнения групп по первичным и вторичным результатам

Пример: «Первичной конечной точкой было изменение массы тела в течение 20 недель исследования в группе пациентов, получавших лечение. Вторичные конечные точки эффективности включали изменение окружности талии, систолического и диастолического артериального давления, распространенность метаболического синдрома. Мы использовали ковариационный анализ (ANCOVA) для первичной конечной точки и для вторичных конечных точек окружности талии, артериального давления и результатов, сообщаемых пациентами; это было дополнено анализом повторных измерений. Модель ANCOVA включала лечение, страну и пол в качестве фиксированных эффектов, а массу тела при рандомизации в качестве ковариации. Мы стремились оценить, свидетельствуют ли данные о превосходстве каждой дозы Лираглутида над плацебо (основная цель) и Орлистатом (вторичная цель)» [176].

Объяснение. Существует много различных методов анализа данных, не все из которых могут быть применимы в конкретной ситуации. Важно указать все методы статистической оценки, использованные при каждом анализе, и затем подробно описать их в разделе «Результаты». Принцип, которому необходимо следовать, заключается в следующем: «Описывать статистические методы достаточно подробно, чтобы компетентный/опытный читатель, имеющий доступ к исходным данным, мог проверить представленные результаты» (www.icmje.org). Также важно описать детали статистического анализа, такие как ИТТ-анализ (*Intention-to-treat analysis*) — анализ по назначенному лечению (вставка 6).

Почти все методы статистического анализа позволяют рассчитать размер эффекта лечения, который определяется различием между исходами в группах сравнения. Для оценки эффекта лечения авторы должны привести его доверительный интервал, который указывает диапазон неопределенности для истинного эффекта лечения. Доверительный интервал можно интерпретировать как диапазон значений эффекта лечения, который совместим с наблюдаемыми данными. Принято считать, что 95-процентный

доверительный интервал дает диапазон, включающий в себя истинное значение 95 из 100 аналогичных исследований.

Результаты исследования можно также оценить с точки зрения их статистической значимости. Значение P представляет собой вероятность того, что выявленные различия могли возникнуть случайно при отсутствии реальных различий эффекта вмешательства. Предпочтительнее приводить фактические значения P (например, $P = 0,003$), а не пороговые значения, такие как $P < 0,05$ [48, 177].

Стандартные методы анализа предполагают, что данные являются «независимыми». Для контролируемых исследований это обычно означает, что у каждого участника зарегистрировано одно значение изучаемого исхода. Обработка нескольких наблюдений от одного участника как независимых данных — серьезная ошибка; такие данные получаются, когда исходы оценивают по результатам измерений различных частей тела, что нередко имеет место в стоматологии и ревматологии. При статистическом анализе данных должен учитываться только один результат наблюдения у каждого участника [178, 179], в противном случае необходимо применять сложные статистические расчеты [180]. Некорректный анализ множественных наблюдений был отмечен в 123 (63%) из 196 исследований у больных ревматоидным артритом [108].

Пункт 126. Методы для дополнительного анализа, такие как анализ подгрупп и скорректированный анализ

Примеры: «При сравнении долей больных с ответной реакцией на лечение в группах вмешательства применялся критерий Х-квадрат (χ^2) Mantel-Haenszel, скорректированный с учетом такой переменной стратификации, как использование метотрексата» [103].

«С помощью стратифицированных моделей пропорциональных рисков Кокса были проведены предварительные анализы подгрупп в соответствии с назначением(ями) антиоксидантной терапии, наличием или отсутствием сердечно-сосудистых заболеваний в анамнезе, потреблением фолиевой кислоты с пищей, курением, диабетом, приемом аспирина, гормональной терапией и использованием поливитаминов. В этих анализах использовались базовые оценки воздействия, и они были ограничены участниками с непропущенными данными о подгруппах на исходном уровне» [181].

Объяснение. Как и в случае первичного анализа, следует четко указать метод анализа в подгруппах. Наиболее убедительными являются методы, которые основаны на выявлении доказательств различия в эффекте лечения при сравнении с дополнительными подгруппами (например, между пожилыми и молодыми участниками). Это сравнение известно как тест для выявления взаимодействий [182, 183]. Распространенным, но менее надежным подходом является сравнение значений P , полученных при оценке размера эффекта лечения при отдельном анализе

в каждой подгруппе. Такой подход может привести к ошибочному выводу о более выраженном эффекте лечения (взаимодействии) в подгруппах на основании одного значимого и одного незначимого значения P [184]. Такие выводы имеют высокий процент ложноположительных результатов.

Из-за высокого риска ложных результатов анализ в подгруппах часто не рекомендуется [14, 185]. Сравнение подгрупп постфактум (или *post hoc* — анализ, проводимый после первоначальной статистической обработки данных), вероятно, не будет подтверждено дальнейшими исследованиями. Подобные анализы не вызывают большого доверия.

В некоторых исследованиях дисбаланс групп по характеристикам участников корректируется с помощью той или иной формы множественного регрессионного анализа. Хотя потребность в корректировке при РКИ значительно меньше, чем в эпидемиологических исследованиях, скорректированный анализ может оказаться целесообразным, особенно если одна или несколько прогностических переменных считаются более важными, чем остальные [186]. В идеале скорректированный анализ должен быть указан в протоколе исследования (см. пункт 24). Например, часто рекомендуется корректировка любых переменных стратификации (см. пункт 8b), исходя из принципа, что стратегия анализа должна соответствовать плану. В РКИ решение о применении скорректированного анализа не должно определяться тем, насколько статистически значимы различия исходных характеристик участников в разных группах (см. пункт 16) [183, 187]. Следует указать обоснование применения любого скорректированного анализа и использованных статистических методов.

Авторы должны обосновать выбор переменных, которые были скорректированы, указать, как были обработаны непрерывные переменные, указать, был ли анализ запланирован или решение о его проведении было принято после полученных данных [188]. Обзоры опубликованных статей об исследованиях показывают, что представление результатов о скорректированных анализах не отвечают всем этим требованиям [188–191].

СПИСОК ЛИТЕРАТУРЫ

1. Rennie D. CONSORT revised-improving the reporting of randomized trials. JAMA 2001;285:2006-7.
2. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273:408-12.
3. Moher D. CONSORT: an evolving tool to help improve the quality of reports of randomized controlled trials. Consolidated Standards of Reporting Trials. JAMA 1998;279:1489-91.
4. Kjaergard LL, Villumsen J, Gluud C. Quality of randomised clinical trials affects estimates of intervention efficacy. Rome, Italy: 7th Cochrane Colloquium; 1999.
5. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ 2001;323:42-6.
6. Veldhuyzen van Zanten SJ, Cleary C, Talley NJ, Peterson TC, Nyren O, Bradley LA, et al. Drug treatment of functional dyspepsia:

a systematic analysis of trial methodology with recommendations for design of future trials. *Am J Gastroenterol* 1996;91:660-73.

7. Talley NJ, Owen BK, Boyce P, Paterson K. Psychological treatments for irritable bowel syndrome: a critique of controlled treatment trials. *Am J Gastroenterol* 1996;91:277-83.

8. Adetugbo K, Williams H. How well are randomized controlled trials reported in the dermatology literature? *Arch Dermatol* 2000;136:381-5.

9. Kjaergard LL, Nikolova D, Gluud C. Randomized clinical trials in HEPATOLOGY: predictors of quality. *Hepatology* 1999;30:1134-8.

10. Schor S, Karten I. Statistical evaluation of medical journal manuscripts. *JAMA* 1966;195:1123-8.

11. Gore SM, Jones IG, Rytter EC. Misuse of statistical methods: critical assessment of articles in BMJ from January to March 1976. *BMJ* 1977;1:85-7.

12. Hall JC, Hill D, Watts JM. Misuse of statistical methods in the Australasian surgical literature. *Aust N Z J Surg* 1982;52:541-3.

13. Altman DG. Statistics in medical journals. *Stat Med* 1982;1:59-71.

14. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med* 1987;317:426-32.

15. Altman DG. The scandal of poor medical research. *BMJ* 1994;308: 283-4.

16. Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet* 2005;365:1159-62.

17. Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ* 2010;340. c723.

18. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 1999;319: 670-4.

19. Lai TY, Wong VW, Lam RF, Cheng AC, Lam DS, Leung GM. Quality of reporting of key methodological items of randomized controlled trials in clinical ophthalmic journals. *Ophthalmic Epidemiol* 2007;14:390-8.

20. Moher D, Fortin P, Jadad AR, Juni P, Klassen T, Le LJ, et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996;347:363-6.

21. Junker CA. Adherence to published standards of reporting: a comparison of placebo-controlled trials published in English or German. *JAMA* 1998;280:247-9.

22. Altman DG. Randomisation. *BMJ* 1991;302:1481-2.

23. Schulz KF, Chalmers I, Grimes DA, Altman DG. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *JAMA* 1994;272:125-8.

24. Streptomycin treatment of pulmonary tuberculosis: a Medical Research Council investigation. *BMJ* 1948;2:769-82.

25. Schulz KF. Randomized controlled trials. *Clin Obstet Gynecol* 1998;41:245-56.

26. Greenland S. Randomization, statistics, and causal inference. *Epidemiology* 1990;1:421-9.

27. Armitage P. The role of randomization in clinical trials. *Stat Med* 1982;1:345-52.

28. Kleijnen J, Gotzsche PC, Kunz R, Oxman AD, Chalmers I. So what's so special about randomisation. In: Maynard A, Chalmers I, editors. *Non-random reflections on health services research*. BMJ Books; 1997. p. 93-106.

29. Chalmers I. Assembling comparison groups to assess the effects of health care. *J R Soc Med* 1997;90:379-86.

30. Nicolucci A, Grilli R, Alexanian AA, Apolone G, Torri V, Liberati A. Quality, evolution, and clinical implications of randomized, controlled trials on the treatment of lung cancer. A lost opportunity for meta-analysis. *JAMA* 1989;262:2101-7.

31. Ah-See KW, Molony NC. A qualitative assessment of randomized controlled trials in otolaryngology. *J Laryngol Otol* 1998;112:460-3.

32. Altman DG, Dore CJ. Randomisation and baseline comparisons in clinical trials. *Lancet* 1990;335:149-53.

33. Thornley B, Adams C. Content and quality of 2000 controlled trials in schizophrenia over 50 years. *BMJ* 1998;317:1181-4.

34. DerSimonian R, Charette LJ, McPeck B, Mosteller F. Reporting on methods in clinical trials. *N Engl J Med* 1982;306:1332-7.

35. A proposal for structured reporting of randomized controlled trials. The Standards of Reporting Trials Group. *JAMA* 1994;272:1926-31.

36. Call for comments on a proposal to improve reporting of clinical trials in the biomedical literature. Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature. *Ann Intern Med* 1994;121:894-5.

37. Rennie D. Reporting randomized controlled trials. An experiment and a call for responses from readers. *JAMA* 1995;273:1054-5.

38. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA* 1996;276:637-9.

39. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA* 2006;295:1152-60.

40. Campbell MK, Elbourne DR, Altman DG. CONSORT statement: extension to cluster randomised trials. *BMJ* 2004;328: 702-8.

41. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008;337. a2390.

42. Ioannidis JP, Evans SJ, Gotzsche PC, O'Neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004;141: 781-8.

43. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med* 2008;148: 295-309.

44. Gagnier JJ, Boon H, Rochon P, Moher D, Barnes J, Bombardier C. Reporting randomized, controlled trials of herbal interventions: an elaborated CONSORT statement. *Ann Intern Med* 2006;144:364-7.

45. Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, et al. CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med* 2008;5:e20.

46. Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. Panel on Cost-Effectiveness in Health and Medicine. *JAMA* 1996;276:1339-41.

47. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ* 1996;313:275-83.

48. Lang TA, Secic M. How to report statistics in medicine. Annotated guidelines for authors, editors, and reviewers. ACP 1997.

49. Davidoff F. News from the International Committee of Medical Journal Editors. *Ann Intern Med* 2000;133:229-31.

50. Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust* 2006;185:263-7.

51. Egger M, Juni P, Bartlett C. Value of flow diagrams in reports of randomized controlled trials. *JAMA* 2001;285:1996e9.

52. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med* 2001;134:657-62.

53. Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001;285:1987-91.

54. Moher D, Schulz KF, Altman DG. The CONSORT statement: re-revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;357:1191-4.

55. Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457-65.

56. Al-Marzouki S, Roberts I, Evans S, Marshall T. Selective reporting in clinical trials: analysis of trial protocols accepted by the Lancet. *Lancet* 2008;372:201.
57. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS ONE* 2008;3:e3081.
58. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663-94.
59. Schulz KF, Altman DG, Moher D for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c332.
60. Pocock SJ. Clinical trials: a practical approach. John Wiley; 1983.
61. Meinert CL. Clinical trials: design, conduct and analysis. Oxford University Press; 1986.
62. Friedman LM, Furberg CD, DeMets DL. Fundamentals of clinical trials. 3rd ed. Springer; 1998.
63. Bolliger CT, Zellweger JP, Danielsson T, van Biljon X, Robidou A, Westin A, et al. Smoking reduction with oral nicotine inhalers: double blind, randomised clinical trial of efficacy and safety. *BMJ* 2000;321:329-33.
64. Dickersin K, Manheimer E, Wieland S, Robinson KA, Lefebvre C, McDonald S. Development of the Cochrane Collaboration's CENTRAL Register of controlled clinical trials. *Eval Health Prof* 2002;25:38-64.
65. Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, et al. CONSORT for reporting randomised trials in journal and conference abstracts. *Lancet* 2008;371:281-3.
66. The impact of open access upon public health. *PLoS Med* 2006;3: e252.
67. Harbourn AM, Knecht LS, Humphreys BL. Structured abstracts in MEDLINE, 1989-1991. *Bull Med Libr Assoc* 1995;83:190-5.
68. Harris AH, Standard S, Brunning JL, Casey SL, Goldberg JH, Oliver L, et al. The accuracy of abstracts in psychology journals. *J Psychol* 2002;136:141-8.
69. Pitkin RM, Branagan MA, Burmeister LF. Accuracy of data in abstracts of published research articles. *JAMA* 1999;281:1110-1.
70. Ward LG, Kendrach MG, Price SO. Accuracy of abstracts for original research articles in pharmacy journals. *Ann Pharmacother* 2004;38:1173-7.
71. Gotzsche PC. Believability of relative risks and odds ratios in abstracts: cross sectional study. *BMJ* 2006;333:231-4.
72. Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *JAMA* 2001;285:437-43.
73. Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. More informative abstracts revisited. *Ann Intern Med* 1990;113: 69-76.
74. Taddio A, Pain T, Fassos FF, Boon H, Ilersich AL, Einarson TR. Quality of nonstructured and structured abstracts of original research articles in the British Medical Journal, the Canadian Medical Association Journal and the Journal of the American Medical Association. *CMAJ* 1994;150:1611-5.
75. Wager E, Middleton P. Technical editing of research reports in biomedical journals. *Cochrane Database Syst Rev* 2008;. MR000002.
76. Hartley J, Sydes M, Blurton A. Obtaining information accurately and quickly: Are structured abstracts more efficient? *J Inform Sci* 1996;22:349-56.
77. Gilligan D, Nicolson M, Smith I, Groen H, Dalesio O, Goldstraw P, et al. Preoperative chemotherapy in patients with resectable non-small cell lung cancer: results of the MRC LU22/NVALT 2/EORTC 08012 multicentre randomised trial and update of systematic review. *Lancet* 2007;369:1929-37.
78. Sandler AD, Sutton KA, DeWeese J, Girardi MA, Sheppard V, Bodfish JW. Lack of benefit of a single dose of synthetic human secretin in the treatment of autism and pervasive developmental disorder. *N Engl J Med* 1999;341:1801-6.
79. World Medical Association. Declaration of Helsinki: ethical principle for medical research involving human subjects. 59th WMA General Assembly, Seoul 2008; www.wma.net/e/policy/b3.htm (accessed 2 June 2009).
80. Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med* 1992;327:248-54.
81. Fergusson D, Glass KC, Hutton B, Shapiro S. Randomized controlled trials of aprotinin in cardiac surgery: could clinical equipoise have stopped the bleeding? *Clin Trials* 2005;2:218-29.
82. Savulescu J, Chalmers I, Blunt J. Are research ethics committees behaving unethically? Some suggestions for improving performance and accountability. *BMJ* 1996;313:1390-3.
83. Sadler LC, Davison T, McCowan LM. A randomised controlled trial and meta-analysis of active management of labour. *BJOG* 2000;107: 909-15.
84. Bath FJ, Owen VE, Bath PM. Quality of full and final publications reporting acute stroke trials: a systematic review. *Stroke* 1998;29: 2203-10.
85. Blumer JL, Findling RL, Shih WJ, Soubrane C, Reed MD. Controlled clinical trial of zolpidem for the treatment of insomnia associated with attention-deficit/hyperactivity disorder in children 6 to 17 years of age. *Pediatrics* 2009;123:e770-6.
86. Sabatine MS, Antman EM, Widimsky P, Ebrahim IO, Kiss RG, Saaiman A, et al. Otamixaban for the treatment of patients with non-ST-elevation acute coronary syndromes (SEPIA-ACS1 TIMI 42): a randomised, double-blind, active-controlled, phase 2 trial. *Lancet* 2009;374:787-95.
87. Grant AM, Altman DG, Babiker AB, Campbell MK, Clemens FJ, Darbyshire JH, et al. Issues in data monitoring and interim analysis of trials. *Health Technol Assess* 2005;9:1-iv.
88. Gallo P, Krams M. PhRMA Working Group on adaptive designs, "White Paper." *Drug Information Journal* 2006;40:421-82.
89. Brown CH, Ten Have TR, Jo B, Dagne G, Wyman PA, Muthen B, et al. Adaptive designs for randomized trials in public health. *Annu Rev Public Health* 2009;30:1-25.
90. Kelly PJ, Sooriyachchi MR, Stallard N, Todd S. A practical comparison of group-sequential and adaptive designs. *J Biopharm Stat* 2005;15:719-38.
91. Pildal J, Chan AW, Hrobjartsson A, Forfang E, Altman DG, Gotzsche PC. Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study. *BMJ* 2005;330:1049.
92. Chan AW, Hrobjartsson A, Jorgensen KJ, Gotzsche PC, Altman DG. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. *BMJ* 2008;337:a2299.
93. Ndekha MJ, van Oosterhout JJ, Zijlstra EE, Manary M, Saloojee H, Manary MJ. Supplementary feeding with either ready-to-use fortified spread or corn-soy blend in wasted adults starting antiretroviral therapy in Malawi: randomised, investigator blinded, controlled trial. *BMJ* 2009;338:1867-75.
94. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?" *Lancet* 2005;365: 82-93.
95. Fuks A, Weijer C, Freedman B, Shapiro S, Skrutkowska M, Riaz A. A study in contrasts: eligibility criteria in a twenty-year sample of NSABP and POG clinical trials. National Surgical Adjuvant Breast and Bowel Program. Pediatric Oncology Group. *J Clin Epidemiol* 1998;51:69-79.
96. Shapiro SH, Weijer C, Freedman B. Reporting the study populations of clinical trials. Clear transmission or static on the line? *J Clin Epidemiol* 2000;53:973-9.
97. Gandhi M, Ameli N, Bacchetti P, Sharp GB, French AL, Young M, et al. Eligibility criteria for HIV clinical trials and generalizability of results: the gap between published reports and study protocols. *AIDS* 2005;19:1885-96.
98. Hall JC, Mills B, Nguyen H, Hall JL. Methodologic standards in surgical trials. *Surgery* 1996;119:466-72.
99. Weiss NS, Koepsell TD, Psaty BM. Generalizability of the results of randomized trials. *Arch Intern Med* 2008;168:133-5.
100. Devereaux PJ, Yang H, Yusuf S, Guyatt G, Leslie K, Villar JC, et al. Effects of extended-release metoprolol succinate in

patients undergoing non-cardiac surgery (POISE trial): a randomised controlled trial. *Lancet* 2008;371:1839-47.

101. Rannou F, Dimet J, Boutron I, Baron G, Fayad F, Mace Y, et al. Splint for base-of-thumb osteoarthritis: a randomized trial. *Ann Intern Med* 2009;150:661-9.

102. Glasziou P, Meats E, Heneghan C, Shepperd S. What is missing from descriptions of treatment in trials and reviews? *BMJ* 2008;336:1472-4.

103. Mease PJ, Goffe BS, Metz J, VanderStoep A, Finck B, Burge DJ. Etanercept in the treatment of psoriatic arthritis and psoriasis: a randomised trial. *Lancet* 2000;356:385-90.

104. McDowell I, Newell C. Measuring health: a guide to rating scales and questionnaires. 3rd ed. New York: Oxford University Press; 2006.

105. Streiner D, Norman C. Health measurement scales: a practical guide to their development and use. 3rd ed. Oxford: Oxford University Press; 2003.

106. Clarke M. Standardising outcomes for clinical trials and systematic reviews. *Trials* 2007;8:39.

107. Sanders C, Egger M, Donovan J, Tallon D, Frankel S. Reporting on quality of life in randomised controlled trials: bibliographic study. *BMJ* 1998;317:1191-4.

108. Gotzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Control Clin Trials* 1989;10:31-56.

109. Marshall M, Lockwood A, Bradley C, Adams C, Joy C, Fenton M. Unpublished rating scales: a major source of bias in randomised controlled trials of treatments for schizophrenia. *Br J Psychiatry* 2000;176:249-52.

110. Jadad AR, Boyle M, Cunnigham C, Kim M, Schachar R. Treatment of Attention-Deficit/Hyperactivity Disorder. Evidence Report/Technology Assessment No. 11. Rockville, MD: U.S. Department of Health and Human Services, Public Health Service, Agency for Healthcare Research and Quality. AHRQ publication no. 00-E005; 1999.

111. Schachter HM, Pham B, King J, Langford S, Moher D. The efficacy and safety of methylphenidate in attention deficit disorder: A systematic review and meta-analysis. Prepared for the Therapeutics Initiative, Vancouver, B.C., and the British Columbia Ministry for Children and Families, 2000.

112. Dargie HJ. Effect of carvedilol on outcome after myocardial infarction in patients with left-ventricular dysfunction: the CAPRICORN randomised trial. *Lancet* 2001;357:1385-90.

113. Chan AW, Krlaza-Jeric K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *CMAJ* 2004;171:735-40.

114. Vermeulen H, Hofland J, Legemate DA, Ubbink DT. Intravenous fluid restriction after major abdominal surgery: a randomized blinded clinical trial. *Trials* 2009;10:50.

115. Fox K, Ford I, Steg PG, Tendera M, Ferrari R. Ivabradine for patients with stable coronary artery disease and left ventricular systolic dysfunction (BEAUTIFUL): a randomised, double-blind, placebo-controlled trial. *Lancet* 2008;372:807-16.

116. Campbell MJ, Julious SA, Altman DG. Estimating sample sizes for binary, ordered categorical, and continuous outcomes in two group comparisons. *BMJ* 1995;311:1145-8.

117. Guyatt GH, Mills EJ, Elbourne D. In the era of systematic reviews, does the size of an individual trial still matter. *PLoS Med* 2008;5:e4.

118. Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. *Lancet* 2005;365:1348-53.

119. Halpern SD, Karlawish JH, Berlin JA. The continuing unethical conduct of underpowered clinical trials. *JAMA* 2002;288:358-62.

120. Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995;311:485.

121. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA* 1994;272:122-4.

122. Freiman JA, Chalmers TC, Smith H Jr, Kuebler RR. The importance of beta, the type II error and sample size in the design and

interpretation of the randomized control trial. Survey of 71 "negative" trials. *N Engl J Med* 1978;299:690-4.

123. Charles P, Giraudeau B, Dechartres A, Baron G, Ravaud P. Reporting of sample size calculation in randomised controlled trials: review. *BMJ* 2009;338:b1732.

124. Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med* 1984;3:409-22.

125. Goodman SN, Berlin JA. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann Intern Med* 1994;121:200-6.

126. Galgiani JN, Catanzaro A, Cloud GA, Johnson RH, Williams PL, Mirels LF, et al. Comparison of oral fluconazole and itraconazole for progressive, nonmeningeal coccidioidomycosis. A randomized, double-blind trial. *Mycoses Study Group. Ann Intern Med* 2000;133:676-86.

127. Connolly SJ, Pogue J, Hart RG, Hohnloser SH, Pfeffer M, Chrolavicius S, et al. Effect of clopidogrel added to aspirin in patients with atrial fibrillation. *N Engl J Med* 2009;360:2066-78.

128. Geller NL, Pocock SJ. Interim analyses in randomized clinical trials: ramifications and guidelines for practitioners. *Biometrics* 1987;43:213-23.

129. Berry DA. Interim analyses in clinical trials: classical vs. Bayesian approaches. *Stat Med* 1985;4:521-6.

130. Pocock SJ. When to stop a clinical trial. *BMJ* 1992;305:235-40.

131. DeMets DL, Pocock SJ, Julian DG. The agonising negative trend in monitoring of clinical trials. *Lancet* 1999;354:1983-8.

132. Buyse M. Interim analyses, stopping rules and data monitoring in clinical trials in Europe. *Stat Med* 1993;12:509-20.

133. Sydes MR, Altman DG, Babiker AB, Parmar MK, Spiegelhalter DJ. Reported use of data monitoring committees in the main published reports of randomized controlled trials: a cross-sectional study. *Clin Trials* 2004;1:48-59.

134. Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA* 2005;294:2203-9.

135. Coutinho IC, Ramos de Amorim MM, Katz L, Bandeira de Ferraz AA. Uterine exteriorization compared with in situ repair at cesarean delivery: a randomized controlled trial. *Obstet Gynecol* 2008;111:639-47.

136. Juni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. In: Egger M, Davey Smith G, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. *BMJ Books*; 2001.

137. Creinin MD, Meyn LA, Borgatta L, Barnhart K, Jensen J, Burke AE, et al. Multicenter comparison of the contraceptive ring and patch: a randomized controlled trial. *Obstet Gynecol* 2008;111:267-77.

138. Tate DF, Jackvony EH, Wing RR. Effects of internet behavioral counseling on weight loss in adults at risk for type 2 diabetes: a randomized trial. *JAMA* 2003;289:1833-6.

139. Lachin JM. Properties of simple randomization in clinical trials. *Control Clin Trials* 1988;9:312-26.

140. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br J Cancer* 1976;34:585-612.

141. Schulz KF, Grimes DA. The Lancet handbook of essential concepts in clinical research. Elsevier; 2006.

142. Altman DG, Bland JM. How to randomise. *BMJ* 1999;319:703-4.

143. Schulz KF. Subverting randomization in controlled trials. *JAMA* 1995;274:1456-8.

144. Enas GG, Enas NH, Spradlin CT, Wilson MG, Wiltse CG. Baseline comparability in clinical trials: prevention of poststudy anxiety. *Drug Information Journal* 1990;24:541-8.

145. Treasure T, MacRae KD. Minimisation: the platinum standard for trials? Randomisation doesn't guarantee similarity of groups; minimisation does. *BMJ* 1998;317:362-3.

146. Sinei SK, Schulz KF, Lamptey PR, Grimes DA, Mati JK, Rosenthal SM, et al. Preventing IUCD-related pelvic infection: the efficacy of prophylactic doxycycline at insertion. *Br J Obstet Gynaecol* 1990;97:412-9.

147. Radford JA, Landorf KB, Buchbinder R, Cook C. Effectiveness of low-Dye taping for the short-term treatment of plantar heel pain: a randomised trial. *BMC Musculoskelet Disord* 2006;7:64.
148. Chalmers TC, Levin H, Sacks HS, Reitman D, Berrier J, Nagalingam R. Meta-analysis of clinical trials as a scientific discipline. I: Control of bias and comparison with large co-operative trials. *Stat Med* 1987;6:315-28.
149. Pocock SJ. Statistical aspects of clinical trial design. *Statistician* 1982;31:1-18.
150. Haag U. Technologies for automating randomized treatment assignment in clinical trials. *Drug Information Journal* 1998;32:11.
151. Piaggio G, Elbourne D, Schulz KF, Villar J, Pinol AP, Gulmezoglu AM. The reporting of methods for reducing and detecting bias: an example from the WHO Misoprostol Third Stage of Labour equivalence randomised controlled trial. *BMC Med Res Methodol* 2003;3:19.
152. Pildal J, Hrobjartsson A, Jorgensen KJ, Hilden J, Altman DG, Gotzsche PC. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *Int J Epidemiol* 2007;36:847-57.
153. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008;336:601-5.
154. McCandlish R, Bowler U, van Asten H, Berridge G, Winter C, Sames L, et al. A randomised controlled trial of care of the perineum during second stage of normal labour. *Br J Obstet Gynaecol* 1998;105:1262-72.
155. Webster J, Clarke S, Paterson D, Hutton A, van Dyk S, Gale C, et al. Routine care of peripheral intravenous catheters versus clinically indicated replacement: randomised controlled trial. *BMJ* 2008;337: a339.
156. Smith SA, Shah ND, Bryant SC, Christianson TJ, Bjornsen SS, Giesler PD, et al. Chronic care model and shared care in diabetes: randomized trial of an electronic decision support system. *Mayo Clin Proc* 2008;83:747-57.
157. Sacks FM, Bray GA, Carey VJ, Smith SR, Ryan DH, Anton SD, et al. Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates. *N Engl J Med* 2009;360: 859-73.
158. Kaptchuk TJ. Intentional ignorance: a history of blind assessment and placebo controls in medicine. *Bull Hist Med* 1998;72:389-433.
159. Guyatt GH, Pugsley SO, Sullivan MJ, Thompson PJ, Berman L, Jones NL, et al. Effect of encouragement on walking test performance. *Thorax* 1984;39:818-22.
160. Gotzsche PC. Blinding during data analysis and writing of manuscripts. *Control Clin Trials* 1996;17:285-90.
161. Karlowski TR, Chalmers TC, Frenkel LD, Kapikian AZ, Lewis TL, Lynch JM. Ascorbic acid for the common cold. A prophylactic and therapeutic trial. *JAMA* 1975;231:1038-42.
162. Noseworthy JH, Ebers GC, Vandervoort MK, Farquhar RE, Yetisir E, Roberts R. The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial. *Neurology* 1994;44:16-20.
163. Carley SD, Libetta C, Flavin B, Butler J, Tong N, Sammy I. An open prospective randomised trial to reduce the pain of blood glucose testing: ear versus thumb. *BMJ* 2000;321:20.
164. Schulz KF, Chalmers I, Altman DG. The landscape and lexicon of blinding in randomized trials. *Ann Intern Med* 2002;136:254-9.
165. Day SJ, Altman DG. Statistics notes: blinding in clinical trials and other studies. *BMJ* 2000;321:504.
166. Montori VM, Bhandari M, Devereaux PJ, Manns BJ, Ghali WA, Guyatt GH. In the dark: the reporting of blinding status in randomized controlled trials. *J Clin Epidemiol* 2002;55:787-90.
167. Cheng K, Smyth RL, Motley J, O'Hea U, Ashby D. Randomized controlled trials in cystic fibrosis (1966-1997) categorized by time, design, and intervention. *Pediatr Pulmonol* 2000;29:1-7.
168. Lang T. Masking or blinding? An unscientific survey of mostly medical journal editors on the great debate. *Med Gen Med* 2000;2:E25.
169. Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Montori VM, et al. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA* 2001;285:2000-3.
170. Haahr MT, Hrobjartsson A. Who is blinded in randomized clinical trials? A study of 200 trials and a survey of authors. *Clin Trials* 2006;3:360-5.
171. Meinert CL. Masked monitoring in clinical trials: blind stupidity? *N Engl J Med* 1998;338:1381-2.
172. Mills E, Prousky J, Raskin G, Gagnier J, Rachlis B, Montori VM, et al. The safety of over-the-counter niacin. A randomized placebo-controlled trial [ISRCTN18054903]. *BMC Clin Pharmacol* 2003;3:4.
173. Schulz KF, Grimes DA, Altman DG, Hayes RJ. Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. *BMJ* 1996;312:742-4.
174. Fergusson D, Glass KC, Waring D, Shapiro S. Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials. *BMJ* 2004;328:432.
175. Sackett DL. Turning a blind eye: why we don't test for blindness at the end of our trials. *BMJ* 2004;328:1136.
176. Astrup A, Rossner S, Van Gaal L, Rissanen A, Niskanen L, Al HM, et al. Effects of liraglutide in the treatment of obesity: a randomised, double-blind, placebo-controlled study. *Lancet* 2009;374:1606-16.
177. Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. In: Altman DG, Machin D, Bryant TN, Gardner MJ, editors. *Statistics with confidence: confidence intervals and statistical guidelines*. 2nd ed. *BMJ Books*; 2000. p. 171-90.
178. Altman DG, Bland JM. Statistics notes. Units of analysis. *BMJ* 1997;314:1874.
179. Bolton S. Independence and statistical inference in clinical trial designs: a tutorial review. *J Clin Pharmacol* 1998;38:408-12.
180. Greenland S. Principles of multilevel modelling. *Int J Epidemiol* 2000;29:158-67.
181. Albert CM, Cook NR, Gaziano JM, Zaharris E, MacFadyen J, Danielson E, et al. Effect of folic acid and B vitamins on risk of cardiovascular events and total mortality among women at high risk for cardiovascular disease: a randomized trial. *JAMA* 2008;299: 2027-36.
182. Matthews JN, Altman DG. Interaction 3: How to examine heterogeneity. *BMJ* 1996;313:862.
183. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 2000;355:1064-9.
184. Matthews JN, Altman DG. Statistics notes. Interaction 2: Compare effect sizes not P values. *BMJ* 1996;313:808.
185. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med* 1992;116:78-84.
186. Steyerberg EW, Bossuyt PM, Lee KL. Clinical trials in acute myocardial infarction: should we adjust for baseline characteristics? *Am Heart J* 2000;139:745-51.
187. Altman DG. Adjustment for covariate imbalance. In: Armitage P, Colton T, editors. *Encyclopedia of biostatistics*. John Wiley; 1998. p. 1000-5.
188. Mullner M, Matthews H, Altman DG. Reporting on statistical methods to adjust for confounding: a cross-sectional survey. *Ann Intern Med* 2002;136:122-6.
189. Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. *Ann Intern Med* 1993;118:201-10.
190. Bender R, Grouven U. Logistic regression models used in medical research are poorly presented. *BMJ* 1996;313:628.
191. Khan KS, Chien PF, Dwarakanath LS. Logistic regression models in obstetrics and gynecology literature. *Obstet Gynecol* 1999;93: 1014-20.