

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

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 прошлом номере мы опубликовали первую часть CONSORT — Консолидированные (общие) стандарты предоставления результатов о рандомизированных клинических исследованиях (РКИ). Во втором номере журнала мы публикуем вторую, завершающую, часть материалов CONSORT. Данный документ предназначен для более широкого применения, понимания и распространения руководства CONSORT, которое должно помочь в улучшении составления статей рандомизированных исследований. Руководство CONSORT 2010 как разъясняющий и уточняющий документ, а также соответствующий сайт (www.consort-statement.org) должны помочь в улучшении составления статей рандомизированных исследований.

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 II. *Проблемы геронауки*. 2024; 2(6): 92–111.

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

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

In our previous issue, we featured the first part of CONSORT — the Consolidated Standards for Reporting Trials (RCTs). Now, in this second issue of our journal, we are pleased to present the second and final part of the CONSORT materials. This extensive document is designed to be widely utilized by individuals seeking to understand and disseminate the CONSORT guidelines. Its primary objective is to enhance the quality

of reports on randomized trials by offering a clear and comprehensive guide. In addition, the CONSORT 2010 is supported by a corresponding website at www.consort-statement.org. Both the CONSORT and the website serve as valuable resources, providing explanations and clarifications to aid in the compilation of articles on randomized trials.

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 II. *Problems of Geroscience*. 2024; 2(6): 92–111.

РЕЗУЛЬТАТЫ

Пункт 13. Поток участников (настоятельно рекомендуется использовать диаграмму)
Пункт 13а. Количество участников при случайном распределении на каждом этапе исследования,

получивших предполагаемое лечение; оценка первичных результатов
Примеры (см. рис. 2, 3).

Рисунок 2.

Блок-схема многоцентрового исследования фракционного резерва кровотока в сравнении с ангиографией при проведении чрескожного коронарного вмешательства (ЧКВ) (адаптировано из Tonino et al [1]). На диаграмме представлена подробная информация об исключенных участниках

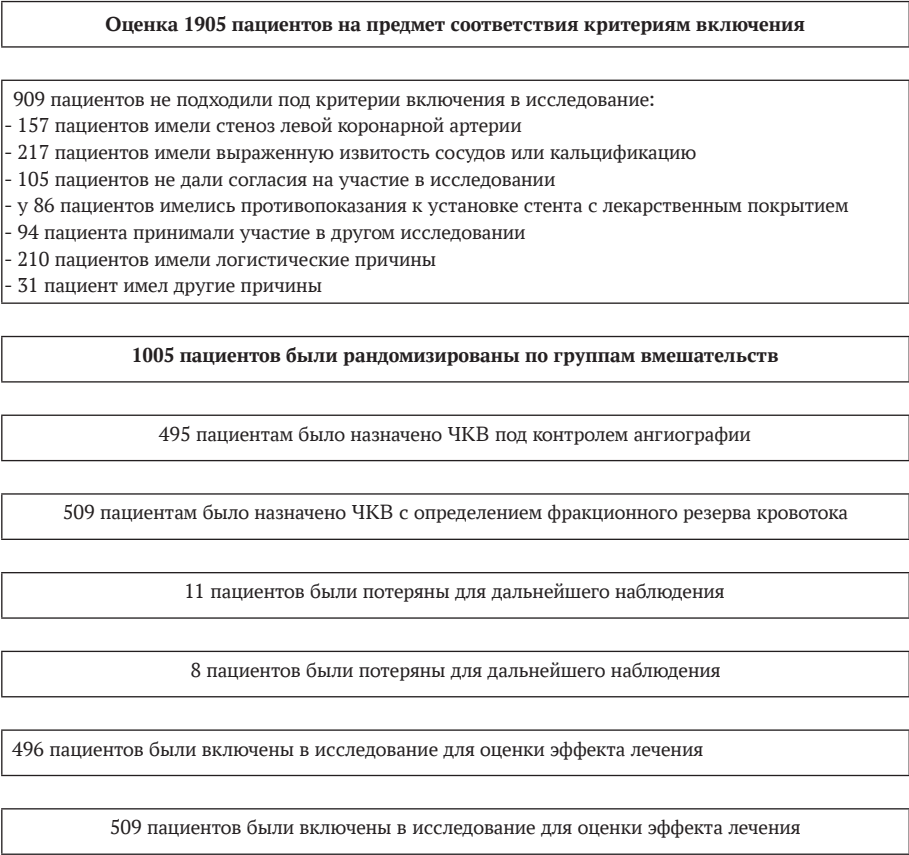
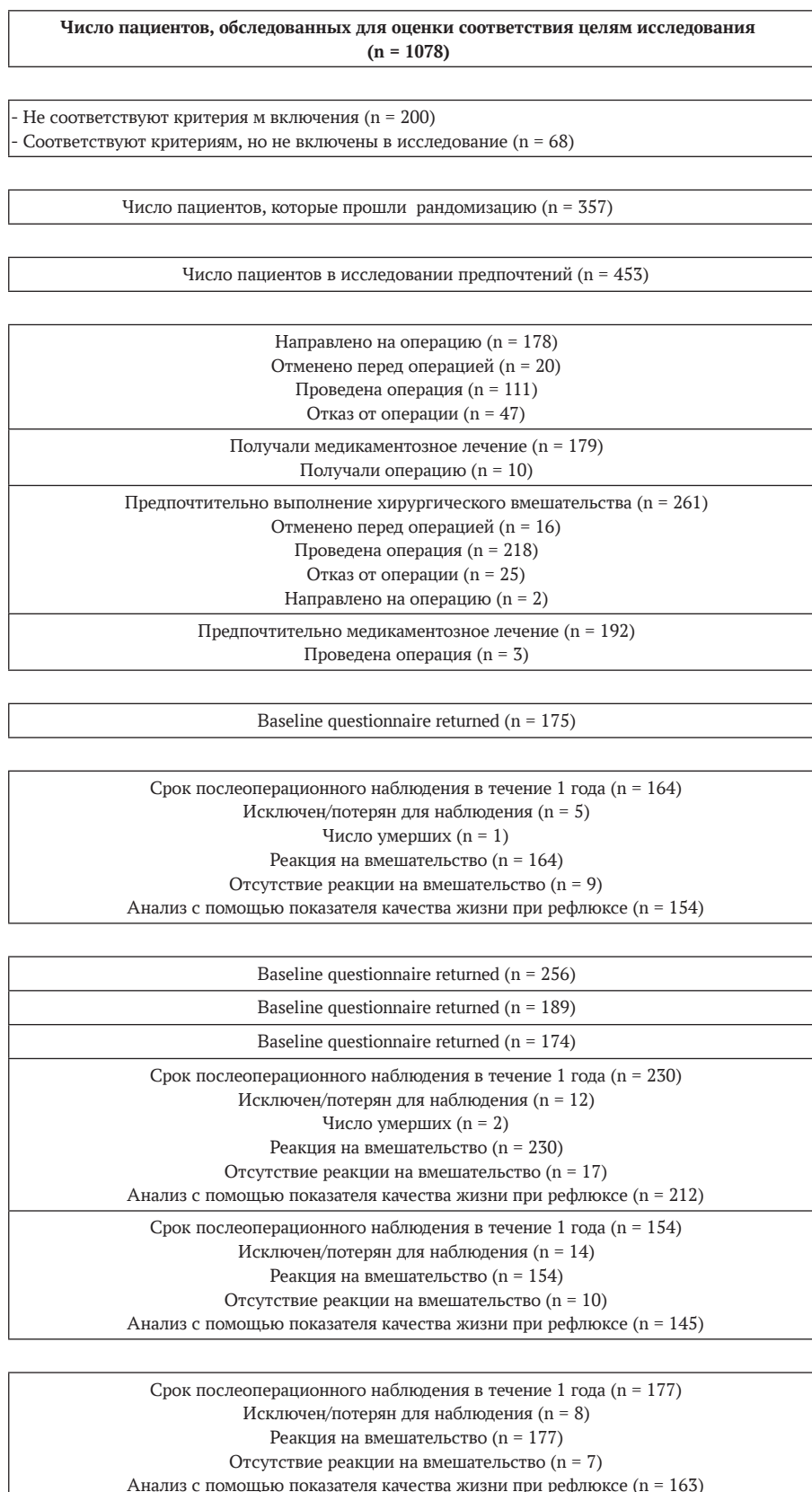


Рисунок 3.

Блок-схема минимального хирургического вмешательства в сравнении с медикаментозным лечением хронической гастроэзофагеальной рефлюксной болезни (адаптировано из Grant et al [2]). На диаграмме показано многоцентровое исследование с параллельной нерандомизированной группой предпочтений



Объяснение. Когда структура и проведение РКИ достаточно просты, в частности если нет потерь до последующего наблюдения или исключений, динамика числа участников на каждом этапе исследования может быть описана в нескольких предложениях. В более сложных исследованиях читателям должно быть разъяснено, почему некоторые участники не получали назначенное лечение, почему некоторые пациенты были из последующего наблюдения или данные о них были исключены из анализа [3]. Эта информация является важной по нескольким причинам. Участники, которые были исключены после распределения, вероятно, имели особенности, нехарактерные в целом для участников исследования. Например, пациенты могут выбыть из дальнейшего наблюдения из-за резкого обострения их заболевания или возникновения тяжелых побочных эффектов от лечения [4, 5].

Отсев участников в результате потери контакта с ними, зачастую являющийся неизбежным, следует отличать от исключения участников из определенного исследования по таким причинам, как несоответствие целям исследования, отказ от лечения и низкая приверженность протоколу исследования. Исключение участников из анализа может приводить к неправильным выводам, особенно вероятна систематическая ошибка при дисбалансе случаев исключения участников в основной и контрольной группах [5-7]. Таким образом, информация о том, включили ли исследователи в анализ всех участников, прошедших рандомизацию в группах, в которые они были первоначально распределены, очень важен (анализ *intention-to-treat* — анализ, проводимый исходя из допущения, что все больные получили назначенное вмешательство, см. в пункте 16 и вставке 6). Знание количества участников, не получивших вмешательства, как было запланировано, или не прошедших полное лечение, позволяет читателю оценить, в какой степени предполагаемая эффективность терапии может быть недооценена по сравнению с размером эффекта в идеальных условиях.

Если возможно, то о количестве больных, которые были обследованы для оценки соответствия целям исследования, также следует сообщить. Хотя этот показатель имеет отношение только к внешней достоверности и, возможно, менее важен, чем другие значения [8], но он является полезным индикатором репрезентативности выборки.

Обзор статей РКИ, опубликованных в 1998 году в пяти ведущих общемедицинских журналах и журналах по внутренним болезням, показал, что представление информации о динамике числа участников часто не соответствует требованиям, особенно это касается числа участников, получивших назначенное лечение, и числа больных, выбывших из наблюдения [3]. Даже совершенно необходимая информация о количестве участников, прошедших рандомизацию, и числе пациентов, исключенных из анализа, отсутствовала в 20% статей [3]. Данная информация была значительно более подробно изложена в статьях, в которых

приводилась рекомендуемая CONSORT схема проведения исследования. Это исследование послужило основой для разработки блок-схемы в пересмотренном заявлении CONSORT [9-11]. Предлагаемый шаблон блок-схемы представлен на рисунке 1, а необходимые подробности описаны в таблице 3.

Некоторая информация, такая как количество лиц, обследованных для оценки соответствия целям испытания, не всегда может быть известна [12], и, в зависимости от характера исследования, некоторые показатели могут быть более значимыми, чем другие. Иногда целесообразно или даже необходимо адаптировать структуру представленного шаблона схемы для конкретного исследования. В некоторых ситуациях может быть полезно добавить другую информацию. Например, блок-схема параллельного группового исследования минимального хирургического вмешательства по сравнению с медикаментозным лечением хронической гастроэзофагеальной рефлюксной болезни также включала параллельную нерандомизированную группу предпочтений (рисунок 3) [2].

Точная форма и содержание блок-схемы могут варьироваться в зависимости от особенностей исследования. Например, многие исследования хирургического вмешательства или вакцинации не предусматривают возможности прекращения лечения. Хотя группа CONSORT настоятельно рекомендует использовать эту схему исследования для отображения потока участников на протяжении всего исследования, однако конкретное предписанное формата не существует.

Пункт 13б. Для каждой группы указаны потери и исключения участников после рандомизации с причинами

Примеры: «Было только одно отклонение от протокола — у женщины из исследуемой группы. Роженице с аномальными размерами таза было назначено плановое кесарево сечение. Однако акушер принял решение о попытке проведения естественных родов, кесарево сечение было сделано лишь при отсутствии прогресса в первом периоде родов» [13].

«В ходе контроля за проведением исследования было закрыто девять центров, в которых отсутствовали доказательства, подтверждающие наблюдение ряда больных или имелись другие серьезные нарушения принципов надлежащей практики клинических исследований» [14].

Объяснение. Некоторые отклонения от протокола исследования могут быть отражены в блок-схеме (см. пункт 13а), например об участниках, которые не получили запланированного вмешательства. Если участники были исключены из исследования после рандомизации (с нарушением принципа *intention-to-treat* — принципа назначения лечения) из-за того, что они не соответствовали критериям включения (см. пункт 16), информацию о них следует включать в блок-схему. Использование термина «нарушение протокола» в опубликованных статьях недостаточно для обоснования исключения участников после

Таблица 3.

Информация, необходимая для документирования динамики числа участников на каждом этапе рандомизированного контролируемого исследования

Этапы РКИ	Число включенных участников	Число пациентов, не включенных или исключенных	Объяснение
Включение	Число пациентов, обследованных для оценки соответствия целям РКИ	Число пациентов, не соответствующих критериям включения в РКИ или соответствующих критериям включения, но отказавшихся от участия в РКИ	Эти показатели позволяют судить о репрезентативности выборки — соответствии характеристик участников РКИ в генеральной совокупности; они используются только для оценки внешней валидности результатов РКИ, но часто не приводятся в статье
Рандомизация	Число пациентов, в случайном порядке отнесенных к каждой выделенной группе		Важные показатели, определяющие размер исследования и позволяющие оценить применение анализа, проводимого исходя из допущения, что все пациенты получили назначенное вмешательство
Вмешательство	Число пациентов в каждой группе, получивших назначенное лечение	Число пациентов в каждой группе, не получивших назначенное лечение	Важные показатели для оценки внутренней валидности и интерпретации результатов РКИ; следует указать причины, по которым отдельные пациенты не получили назначенного лечения
Наблюдение	Число пациентов в каждой группе, которые получили назначенное вмешательство в полном объеме Число пациентов в каждой группе, которые наблюдались до конца исследования	Число пациентов в каждой группе, которые не получили назначенное вмешательство в полном объеме Число пациентов в каждой группе, которые не наблюдались до конца исследования	Важные показатели для оценки внутренней валидности и интерпретации результатов РКИ; следует указать причины, по которым отдельные пациенты не получили назначенное лечение в полном объеме или не наблюдались до конца исследования
Анализ данных	Число пациентов в каждой группе, данные о которых были включены в основной анализ	Число пациентов в каждой группе, данные о которых были исключены из основного анализа	Важный показатель для оценки анализа, проводимого исходя из допущения, что все пациенты получили необходимое лечение; следует указать причины исключения участника из анализа

рандомизации. Всегда следует сообщать о характере отклонения от протокола и точной причине исключения участников после рандомизации.

Пункт 14а. Сроки, определяющие периоды включения участников и последующего наблюдения

Пример: «Включение участников, соответствующих критериям возраста, проводилось с февраля 1993 года по сентябрь 1994 года... Участники посещали клинику во время рандомизации (исходный уровень), а затем через каждые 6 месяцев в течение 3 лет» [15].

Объяснение. Читатель должен знать, когда проводилось исследование и в течение какого периода были набраны участники. Методы медикаментозной терапии и хирургическое лечение, включая сопутствующие методы лечения, постоянно развиваются (совершенствуются) и могут повлиять на стандартную терапию, оказываемую участникам во время исследования. Информация о скорости включения участников также может быть полезной, особенно для других исследователей.

Продолжительность наблюдения не всегда является фиксированным периодом после рандомизации. Во многих РКИ, результатом которых является время наступления события, наблюдение за всеми участниками заканчивается после регистрации определенного

исхода. Необходимо указать эту дату в статье, а также отметить минимальную, максимальную и среднюю продолжительность наблюдения [16, 17].

Обзор статей об исследованиях, опубликованных в журналах по онкологии, в которых использовался анализ выживаемости и большинство из которых не являлись РКИ [17], показал, что почти в 80% (104 из 132) статей были указаны даты начала и окончания включения больных, но лишь в 24% (32 из 132) статей сообщалось о дате окончания периода наблюдения.

Пункт 14б. Причины прекращения или перерыва исследования

Примеры: «На момент промежуточного анализа общий объем наблюдения включал примерно 63% от общего количества пациенто-лет, которое должно было быть собрано в конце исследования, что привело к пороговому значению 0,0095, как определено по методу функции альфа-расходов Лан-ДеМеца... При промежуточном анализе RR составил 0,37 в группе вмешательства по сравнению с контрольной группой, со значением $P = 0,00073$, что ниже порогового значения. Совет по мониторингу данных и безопасности посоветовал прервать исследование и предложить обрезание контрольной группе, которую затем попросили прийти в исследовательский центр, где

было рекомендовано и предложено МС (медицинское обрезание)... Поскольку исследование было прервано, некоторые участники не имели полного наблюдения на тот день, и их визиты, которые еще не были завершены, описаны в этой статье как «запланированные» [18].

«В январе 2000 года проблемы с поставками вакцин привели к необходимости временной общенациональной замены цельноклеточного компонента комбинированной вакцины DPT/Hib на бесклеточную коклюшную вакцину. Поскольку эта вакцина имеет другой профиль местной реактогенности, мы решили досрочно прекратить исследование» [19].

Объяснение. Можно утверждать, что исследователи, которые произвольно проводят незапланированный промежуточный анализ после очень небольшого количества событий, не используя статистические рекомендации, подвергаются высокому риску «выявить» данные в случайном порядке, что, вероятно, будет указывать на выраженное переоценивание пользы от лечения [20].

Читатели, вероятно, сделают более слабые выводы из исследования, которое было сокращено на основе данных, по сравнению с исследованием, в котором результаты сообщаются после достижения цели независимо от полученных результатов. Таким образом, РКИ должны указать, почему исследование закончилось (см. вставку 5). В статье также должны быть раскрыты факторы, не относящиеся к исследованию, которые повлияли на решение о прекращении исследования, а также лица, принявшие решение о прекращении исследования, включая сообщение о роли, которую финансирующее агентство сыграло в обсуждениях и в решении остановить исследование [21].

Систематический обзор 143 РКИ, прекращенных раньше, чем планировалось, для получения пользы, показал, что в этих исследованиях сообщали о прекращении после достижения медианы 66 событий, оценивали средний относительный риск 0,47 и сильную связь между количеством накопленных событий и размером эффекта с меньшими исследованиями с меньшим количеством событий, дающими наибольший эффект от лечения (отношение шансов 31, 95%, доверительный интервал от 12 до 82) [21]. В то время как все большее число исследований, опубликованных во влиятельных медицинских журналах, сообщает о досрочном прекращении, только в 0,1% исследований сообщается о досрочном прекращении ради пользы, что контрастирует с оценками, полученными на основе имитационных исследований [22] и опросов комитетов по безопасности данных и мониторингу [23]. Таким образом, многие исследования, в которых было мало участников и в которых сообщалось о значительном эффекте лечения, возможно, были остановлены раньше, чем планировалось, но об этом не сообщалось.

Вставка 5. Раннее прекращение исследования

РКИ могут завершиться, когда достигнут целевого размера выборки, целевого количества событий, продолжительности

последующего наблюдения или когда достигнут запланированной даты окончания исследования. В таких ситуациях исследование будет остановлено независимо от его результатов, и остановка вряд ли приведет к искажению результатов. Альтернативно РКИ могут прекратиться раньше, чем планировалось, из-за получения результатов промежуточного анализа, показывающих большую, чем ожидалось, пользу или вред от экспериментального вмешательства. Кроме того, РКИ могут прекратиться раньше, чем планировалось, если исследователи обнаружат доказательства отсутствия существенной разницы между экспериментальными и контрольными вмешательствами (то есть прекращение из-за неэффективности исследуемого вмешательства). Кроме того, исследования могут прекратиться досрочно, потому что исследование становится нежизнеспособным: финансирование прекращается, исследователи не могут получить доступ к подходящим пациентам или вмешательствам в исследовании, или результаты других исследований ставят вопрос о актуальности исследования.

Полное информирование о том, почему исследование закончилось, важно для принятия решений, основанных на фактических данных (см. пункт 14b). Исследователи, изучившие, почему 143 исследования были прекращены досрочно из-за пользы, обнаружили, что многие из них не предоставили ключевую информацию о том, как было принято решение о прекращении: на основании запланированного размера выборки (n528), результатов промежуточного анализа (n545) [21]. Пункт 7b контрольного списка требует предоставления информации о сроках проведения промежуточных анализов, о том, что их вызвало, сколько их было проведено, были ли они запланированными или специальными, а также о том, существовали ли статистические руководящие принципы и правила остановки. Кроме того, полезно знать, участвовал ли в анализе независимый комитет по мониторингу данных (и кто его составлял, уделяя особое внимание роли источника финансирования) и кто принял решение о прекращении. Часто комитет по безопасности и мониторингу данных дает рекомендации, а спонсоры или исследователи принимают решение о прекращении.

Исследования, которые прекращаются досрочно по причинам, очевидно, не зависящим от результатов исследования, а также исследования, которые достигают запланированного прекращения, вряд ли приведут к предвзятости из-за остановки [24]. В этих случаях авторы должны сообщить, проводился ли промежуточный анализ и были ли эти результаты доступны, предоставив их спонсору.

Стремление к проведению исследований, которые изменяют вмешательство в ответ на промежуточные результаты, что позволяет быстрее оценить перспективные вмешательства при быстро развивающихся и фатальных состояниях, потребует еще более тщательного информирования о процессе и решения о досрочном их прекращении [25]

Пункт 15. Таблица, показывающая исходные демографические и клинические характеристики для каждой группы

Пример: см. таблицу 4.

Объяснение. Хотя описание критериев включения (см. пункт 4a) позволяет судить о том, какие больные соответствуют целям исследования, также важно знать характеристики участников, которые фактически были включены в исследование. Эта информация позволяет читателям, особенно врачам, судить о том, насколько актуальны результаты исследования для конкретного пациента.

Рандомизированные контролируемые исследования направлены на сравнение групп участников,

которые различаются только в отношении применяемого вмешательства (лечения). Хотя правильное случайное распределение предотвращает систематическую ошибку, связанную с отбором участников, оно не гарантирует однородность групп по исходным характеристикам участников исследования. Однако любые различия в исходных характеристиках являются результатом случайности, а не отражают систематическую ошибку [25]. Чтобы оценить, насколько группы исследования сопоставимы, нужно сравнить важные исходные демографические и клинические характеристики их участников. Исходные данные особенно ценны для результатов, которые также можно измерить в начале исследования (например, артериальное давление).

Исходные демографические и клинические характеристики участников удобно представить в виде таблицы (см. таблицу 4). Для непрерывных переменных, таких как масса тела или артериальное давление, наряду со средними значениями следует указывать показатели вариабельности. В каждой группе непрерывные переменные могут быть представлены средним значением и стандартным отклонением. Когда непрерывные данные имеют асимметричное распределение, предпочтительным подходом может быть указание медианы и диапазона в перцентилях (например,

с 25-го по 75-й перцентиль) [26]. Стандартные ошибки и доверительные интервалы не подходят для описания изменчивости данных — они скорее относятся к проверке гипотез, а не к описательной статистике. Переменные с небольшим количеством упорядоченных категорий (например, стадии заболевания от I до IV) не следует рассматривать как непрерывные переменные; вместо этого следует указывать количество участников (в абсолютных числах и долях) в каждой категории [27, 26].

К сожалению, оценка статистической значимости различий исходных характеристик участников все еще распространена [28, 25, 29]; она была зарегистрирована в половине из 50 исследований РКИ, опубликованных в ведущих журналах в 1997 году [28]. Такая оценка может говорить о вероятности того, что наблюдаемые различия в исходных характеристиках могли возникнуть случайно. Тем не менее мы уже знаем, что любые различия вызваны случайно. Оценка исходных различий необязательно ошибочна, просто нелогична [30]. Такая проверка гипотез является излишней и может ввести в заблуждение исследователей и их читателей. Вероятнее, сравнения на исходном уровне должны основываться на рассмотрении прогрессивной силы измеренных переменных и размера любых случайных дисбалансов, которые произошли [30].

Таблица 4.

Пример представления данных об исходных демографических и клинических характеристиках участников в каждой из групп исследования

Характеристика	Телмисартан (N = 2954)	Плацебо (N = 2972)
Возраст (годы)	66,9 (7,3)	66,9 (7,4)
Пол (женский)	1280 (43,3%)	1267 (42,6%)
Статус курения:		
- текущий	293 (9,9%)	289 (9,7%)
- прошлый	1273 (43,1%)	1283 (43,2%)
Этническая принадлежность:		
- азиатская	637 (21,6%)	624 (21,0%)
- арабская	37 (1,3%)	40 (1,3%)
- африканская	51 (1,7%)	55 (1,9%)
- европеоидная	1801 (61,0%)	1820 (61,2%)
- коренная или абориген	390 (13,2%)	393 (13,2%)
- другие	38 (1,3%)	40 (1,3%)
Артериальное давление (мм рт. ст.)	140,7 (16,8/81,8) (10,1)	141,3 (16,4/82,0) (10,2)
Частота сердечных сокращений (удар в минуту)	68,8 (11,5)	68,8 (12,1)
Уровень холестерина (ммоль/л)		
- общий	5,09 (1,18)	5,08 (1,15)
- ЛПНП	3,02 (1,01)	3,03 (1,02)
- ЛПВП	1,27 (0,37)	1,28 (0,41)
Ишемическая болезнь сердца	2211 (74,8%)	2207 (74,3%)
Инфаркт миокарда	1381 (46,8%)	1360 (45,8%)
Стенокардия	1412 (47,8%)	1412 (47,5%)
Поражение периферических артерий	349 (11,8%)	323 (10,9%)
Артериальная гипертензия	2259 (76,5%)	2269 (76,3%)
Сахарный диабет	1059 (35,8%)	1059 (35,6%)

* Данные представляют собой средние значения (SD) или числа (%). Адаптировано из таблицы 1 Yusuf и др. [31].

Пункт 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 по первоначально определенным группам

Примеры: «Первичный анализ был проведен исходя из того, что все пациенты получили лечение в соответствии с распределением в группы, в анализ были включены данные обо всех участниках, прошедших рандомизированное распределение по группам вмешательств» [32].

«Один пациент в группе лечения Алендронатом был исключен из последующего наблюдения, таким образом, данные 31 пациента, которые получили назначенное лечение, были включены для проведения анализа. В пяти случаях были отмечены нарушения протокола исследования, таким образом, осталось 26 пациентов, данные которых в соответствии с протоколом были включены в окончательный анализ» [33].

Объяснение. Количество участников в каждой группе является важным элементом оценки результатов исследования. Хотя схема проведения исследования (см. пункт 13а) может указывать на количество проанализированных участников, эти цифры часто варьируются для разных показателей результатов. Количество участников в группе должно быть указано для всех анализов. Эта информация особенно важна для бинарных результатов, так как такие оценки размера эффекта, как, например, отношение рисков и разность рисков, следует рассматривать в сопоставлении с частотой анализируемого исхода. Представление результатов в виде дробей также помогает читателю оценить, все ли случайным образом распределенные участники были включены в анализ, и если нет, то сколько участников было исключено. Отсюда следует, что результаты не должны быть представлены исключительно в виде итоговых показателей, таких как относительные риски.

Иногда участники могут получать назначенное вмешательство не в полном объеме или может быть ошибочная рандомизация участников, не соответствующих критериям включения. Одним из широко рекомендуемых способов решения таких проблем является анализ всех участников в соответствии с их первоначальным распределением в группы вмешательств, независимо от того, что впоследствии произошло (вставка 6). Однако такой анализ *intention-to-treat* не всегда легко осуществим. Нередко отдельные больные досрочно выбывают из исследования или им приходится отменять назначенное вмешательство, и, следовательно, в конце исследования исходы у этих больных не оцениваются. Если результатом является смерть, то такие пациенты могут быть включены в анализ на основе исходной информации, в то время как методы условного расчета могут потребоваться в случае отсутствия других итоговых данных. Термин *intention to treat analysis* нередко неправомерно используется и в тех случаях, когда до конца исследования наблюдались все участники, но данные о некоторых из них не были включены в анализ, например,

из-за того, что в действительности лечение не было получено в связи с несоблюдением назначений [34].

С другой стороны, анализ может быть ограничен включением только тех участников, которые соответствуют требованиям протокола с точки зрения целей исследования, полноты проведения предписанного лечения, а также оценки исходов. Такой анализ называют анализом в зависимости от реально полученного вмешательства (*on-treatment*) или анализом в соответствии с протоколом (*per protocol*). Исключение участников из анализа данных может привести к ошибочным выводам. Например, в исследовании, в котором сравнивались консервативное и хирургическое лечение при стенозе сонной артерии, анализ, ограниченный данными участников, которые наблюдались до конца исследования, показал, что хирургическое вмешательство снижает риск развития преходящего нарушения мозгового кровообращения, инсульта и смерти. Однако анализ *intention-to-treat*, основанный на включении всех участников, первоначально распределенных в группы вмешательств, не показал преобладающего эффекта хирургического лечения [35].

Анализ *intention-to-treat*, как правило, предпочтительнее, потому что он позволяет избежать систематической ошибки, связанной с неслучайным выбыванием участников из исследования [36–38]. Независимо от того, используют ли авторы термин *intention-to-treat*, они должны пояснить, данные о каких участниках были при этом использованы (см. пункт 13). Несоблюдение назначенной терапии может означать, что анализ *intention to treat* приводит к недооценке реальной эффективности вмешательства; в таких случаях могут потребоваться дополнительные анализы, такие как анализ по протоколу [39, 40]. Однако следует отметить, что такие анализы часто содержат значительные ошибки [41].

В обзоре 403 РКИ, опубликованных в 10 ведущих медицинских журналах в 2002 году, 249 (62%) сообщили об использовании анализа *intention-to-treat* для своего первичного анализа. Эта доля была выше среди журналов, придерживающихся заявления CONSORT (70% против 48%). Среди статей, в которых сообщалось об использовании анализа *intention-to-treat*, только 39% действительно анализировали всех участников как рандомизированных, при этом более 60% статей не содержали данные в их первичном анализе [42]. Другие исследования показывают аналогичные результаты [34, 43, 44]. Исследования, в которых не сообщалось обо всех случаях исключения данных из анализа, были методологически более слабыми и в других отношениях, чем те исследования, в которых указаны все случаи исключения; это наглядно свидетельствует о сокрытии некоторыми авторами важной информации об исключении данных из анализа [45]. Другое исследование показало, что статьи, в которых сообщается о проведении анализа *intention-to-treat*, имеют и другие признаки соответствия добросовестной практике научных исследований и публикаций, в частности в них описывается процедура расчета размера выборки [46].

Вставка 6. Анализ в соответствии с назначенным вмешательством

Особой силой РКИ является отсутствие систематической ошибки при распределении участников исследования по группам вмешательств (вставка 1). Это позволяет получить убедительные выводы о причине и следствии, которые не оправданы другими дизайнами исследования. Чтобы полностью сохранить важность рандомизации, мы должны анализировать данные обо всех участниках в соответствии с тем, к какой группе вмешательства они были отнесены первоначально (вне зависимости от того, применялось ли в действительности это вмешательство). Эти два условия определяют анализ в соответствии с назначенным вмешательством, который широко рекомендуется в качестве предпочтительной стратегии анализа [34, 44]. Трудно достичь строгого анализа intention-to-treat по двум основным причинам — отсутствие результатов для некоторых участников и несоблюдение протокола исследования.

Отсутствие результатов

Многие исследователи исключают пациентов без наблюдаемого результата. Часто это разумно, но как только какие-либо рандомизированные участники исключаются, анализ не является анализом в соответствии с назначенным вмешательством. Действительно, большинство рандомизированных исследований имеют некоторые отсутствующие наблюдения. Исследователи должны эффективно выбирать между пропуском участников без окончательных данных или условным расчетом их недостающих данных о результатах [45]. Анализ «полного случая» (или «доступного случая») включает только тех, чей результат известен. Хотя некоторые пропущенные результаты не вызовут проблем, в половине исследований более 10% рандомизированных пациентов могут не иметь результатов [46]. Этот общий подход потеряет силу за счет уменьшения размера выборки, и смещение может быть внесено в результаты, если потеря для последующего наблюдения связана с реакцией пациента на лечение. Должно вызывать беспокойство, когда частота или причины выбывания различаются между группами вмешательства.

Участники с отсутствующими результатами могут быть включены в анализ, только если их результаты условно определены (то есть их результаты оцениваются из другой информации, которая была собрана). Условный расчет отсутствующих данных позволяет проводить анализ, но требует сильных допущений, которые могут быть трудно обоснованы [47]. Простые методы расчета эффективны, но их использование может быть нецелесообразным. В частности, широко используемым методом является LOCF (Last observation carried forward — повторение результата последнего наблюдения), в котором отсутствующие окончательные данные переменной результата заменяются последним известным значением до того, как участник был потерян для последующего наблюдения. Это эффективно благодаря своей простоте, но метод может привести к предвзятости [157] и не учитывать неопределенность мнения [158]. Многие авторы подвергли резкой критике данный метод [157, 158, 159].

Несоблюдение протокола исследования

Отдельный вопрос заключается в том, что протокол исследования, возможно, не был полностью соблюден для некоторых участников исследования. Типичными примерами являются участники, которые не соответствовали критериям включения (например, неправильный диагноз, молодой возраст для исследования), получали запрещенное вмешательство, не принимали назначенное лечение, получали другое лечение или не прошли обследования. Простой способ справиться с любыми отклонениями протокола — это игнорировать их: все участники могут быть включены в анализ независимо от их приверженности к протоколу, и это подход анализа в соответствии с назначенным вмешательством. Таким образом, исключение любых участников по таким причинам несовместимо с intention-to-treat analysis.

Термин «modified intention to treat» довольно широко используется для описания анализа, который исключает участников, которые неадекватно придерживались протокола, в частности тех, кто не получил определенное минимальное вмешательство [159]. Альтернативный термин — «по протоколу». Хотя анализ «по протоколу» может быть уместным в некоторых условиях, он должен быть правильно отмечен как нерандомизированное сравнение наблюдений. Любое исключение пациентов из анализа ставит под угрозу рандомизацию и может привести к систематической ошибке в результатах. Как и «анализ в соответствии с назначенным вмешательством», ничто надежно не уточняет, какие именно пациенты были включены. Таким образом, в контрольном списке CONSORT мы убрали обязательное указание на intention-to-treat в пользу четкого описания того, кто именно был включен в каждый анализ.

Пункт 17а. Для каждого первичного и вторичного исхода представлены результаты по каждой группе, а также предполагаемый размер эффекта и его точность (например, доверительный интервал 95%)

Примеры: см. таблицы 5 и 6.

Объяснение. Для каждого из оцениваемых исходов исследования в каждой группе должны быть представлены итоговые показатели (например, доли случаев с возникновением или без возникновения исхода или среднее значение и стандартное отклонение измерений), а также различие результатов в группе лечения и контроля. Для бинарных исходов величина эффекта может быть отношением рисков (относительный риск), отношением шансов или разницей рисков; для данных о выживаемости за период времени — как отношение шансов выживания или отношение риска смерти или разность медиан продолжительности жизни; для непрерывных данных — обычно как различия средних показателей. Доверительные интервалы должны быть представлены при сравнении групп. Распространенной ошибкой является представление отдельных доверительных интервалов для результата в каждой группе, а не для их разницы [50]. Результаты исследований часто более четко отображаются в виде таблицы, чем в виде текста, как показано в таблицах 5 и 6.

Для всех результатов авторы должны предоставить доверительный интервал, чтобы указать точность оценки [51, 52]. Обычно используют 95-процентный доверительный интервал, хотя в ряде случаев можно применить другие его диапазоны. Многие журналы требуют или настоятельно рекомендуют использовать доверительные интервалы [53]. Они особенно важны для оценки результатов со статистически несущественными различиями, для которых они часто указывают, что результат не исключает важных клинически значимых различий. Использование доверительных интервалов заметно возросло в последние годы, хотя и не во всех медицинских специальностях [50]. Доверительные интервалы в отчете могут быть дополнены значением Р, но в любом случае результаты не должны представляться только в виде значения Р [54, 55]. Результаты должны сообщаться для всех запланированных первичных и вторичных

конечных точек, а не только для тех анализов, которые были статистически значимыми или «интересными». Выборочная отчетность в рамках исследования является широко распространенной и серьезной проблемой [56, 57]. В исследованиях, в которых выполнялся промежуточный анализ, интерпретация должна быть сосредоточена на окончательных результатах исследования, а не на промежуточных результатах [58].

Как при анализе бинарных исходов, так и при анализе выживаемости результаты можно представить также в виде числа больных, которых необходимо лечить определенным методом в течение определенного времени, чтобы достичь определенного благоприятного исхода или предотвратить определенный неблагоприятный исход (см. пункт 21) [59, 60].

Таблица 5.

Пример представления итоговых результатов анализа для каждой группы исследования (бинарные результаты)*

Критерий оценки (конечная точка)	Группа (%)		Разность рисков (95% ДИ)
	Этанерцепт (n = 30)	Плацебо (n = 30)	
Первичная: Число пациентов, у которых результат лечения через 12 недель соответствовал критериям эффективного лечения псориатического артрита	26 (87)	7 (23)	63% (44–83)
Вторичная: Число пациентов, соответствующих критериям оценки Американской коллегии ревматологов (АКР):			
ACR20	22 (73)	4 (13)	60% (40–80)
ACR50	15 (50)	1 (3)	47% (28–66)
ACR70	4 (13)	0 (0)	13% (1–26)

PsARC5 — критерии ответа на псориатический артрит.

ACR5 — Американский колледж ревматологии.

*См. также пример для пункта 6а.

Адаптировано из таблицы 2 Mease et al [61].

Таблица 6.

Пример отчета об итоговых результатах для каждой группы вмешательства (непрерывные результаты)

	Лечебная физкультура (n = 65)		Группа контроля (n = 66)		Скорректированное значение (95% ДИ) на 12 месяцев
	Исходный уровень (стандартное отклонение)	Уровень через 12 месяцев (стандартное отклонение)	Исходный уровень (стандартное отклонение)	Уровень через 12 месяцев (стандартное отклонение)	
Оценка функции (0-100)	64,4 (13,9)	83,2 (14,8)	65,9 (15,2)	79,8 (17,5)	4,52 (-0,73 — 9,76)
Боль в покое (0-100)	4,14 (2,3)	1,43 (2,2)	4,03 (2,3)	2,61 (2,9)	-1,29 (-2,16 — -0,42)
Боль при движении (0-100)	6,32 (2,2)	2,57 (2,9)	5,97 (2,3)	3,54 (3,38)	-1,19 (-2,22 — -0,16)

*Оценка функции скорректирована с учетом исходного уровня, возраста и продолжительности симптомов.

Адаптировано из таблицы 3 van Linschoten [62]

Пункт 17б. Для бинарных результатов рекомендуется представление как абсолютных, так и относительных величин эффекта

Пример: «Риск кислородной зависимости или смерти снизился на 16% (95% ДИ от 25% до 7%). Абсолютная разница составила -6,3% (95% ДИ от -9,9% до -2,7%); таким образом, раннее введение препарата

примерно 16 детям предотвратило бы смерть 1 ребенка или длительную зависимость от кислорода» (см. таблицу 7) [63].

Объяснение. Когда первичный результат является бинарным, следует указывать как относительный эффект (отношение рисков (относительный риск) или отношение шансов), так и абсолютный эффект

(разницу рисков) (с доверительными интервалами), поскольку ни относительный показатель, ни абсолютный показатель сами по себе не дают полную картину эффективности и его последствий. Разные аудитории могут предпочесть относительный или абсолютный риск, но и врачи, и неспециалисты склонны переоценивать эффект, когда он представлен в терминах относительного риска [64-66]. Величину разницы в риске нельзя обобщать для других групп населения в меньшей степени, чем относительный риск, поскольку она

зависит от исходного риска в группе, не подвергшейся воздействию, которая имеет тенденцию варьироваться в зависимости от группы населения. Для болезней с общим исходом относительный риск, близкий к единице, может указывать на клинически важные различия в условиях общественного здравоохранения. Напротив, большой относительный риск при редком исходе может быть не столь важен для общественного здравоохранения (хотя он может быть важен для человека, относящегося к категории высокого риска).

Таблица 7.

Пример указания об абсолютной и относительной величине эффекта

Первичный результат	Процент (№)		Коэффициент риска (95% ДИ)	Разница рисков (95% ДИ)
	Раннее введение (n = 1344)	Отсроченное селективное введение (n = 1346)		
Смерть или кислородная зависимость на ожидаемую дату родов	31,9 (429)	38,2 (514)	0,84 (0,75–0,93)	-6,3 (-9,9 — -2,7)

Адаптировано из таблицы 3 The OSIRIS Collaborative Group [63].

Пункт 18. Результаты любых других проведенных анализов, включая анализ подгрупп и скорректированный анализ, отличающийся от предварительно определенного анализа

Пример: «На основании результатов исследования, в котором предполагалось, что эффективность периперационных β -блокаторов может зависеть от исходного риска, мы провели анализ первичных подгрупп на основе пересмотренной системы оценки индекса сердечного риска. Мы также провели анализ вторичных подгрупп по полу, типу операции, использованию эпидуральной или спинальной анестезии. Для всех подгрупповых анализов использовались пропорциональные модели риска Кокса, включающие тесты на взаимодействие, которые считались значимыми при $p < 0,05$... На рисунке 3 показаны результаты анализа подгрупп, проведенного в соответствии с предварительной спецификацией, и отмечена согласованность эффектов. Наши подгрупповые анализы не позволили обнаружить различия в эффектах подгрупп, которые можно было бы ожидать при наличии истинного подгруппового эффекта» [67].

Объяснение. Множественные анализы одних и тех же данных существенно повышают риск получения ложноположительных результатов [68]. Авторам особенно не следует поддаваться искушению проведения большого числа анализов в подгруппах [70, 71, 69]. Результаты анализов, запланированных в протоколе РКИ (см. пункт 24), значительно надежнее результатов тех анализов, решение о проведении которых было обусловлено полученными данными. Авторы должны указать, какие анализы были заранее запланированы. Если проводились анализы данных в подгруппах, необходимо описать эти подгруппы и обосновать их выбор, хотя приводить подробные разъяснения не всегда обязательно. Избирательное представление результатов этих анализов может способствовать

возникновению систематической ошибки [73]. Для правильной оценки взаимодействия (см. пункт 126) данные следует представлять в виде различия оценок эффекта вмешательства между каждой из подгрупп (вместе с диапазоном доверительного интервала), не ограничиваясь только значением P .

В одном из исследований 35 из 50 статей об исследованиях включали анализ подгрупп, из которых только 42% использовали тесты взаимодействия [70]. Часто было трудно определить, был ли анализ подгрупп указан в протоколе. В другом обзоре хирургических исследований, опубликованных в журналах с высокой импактностью, 27 из 72 исследований сообщили о 54 подгрупповых анализах, из которых 91% были *post hoc* и только в 6% подгрупповых анализов использовался тест на взаимодействие для оценки наличия эффекта подгруппы [74].

Подобные рекомендации применимы к анализу, скорректированному с учетом исходных характеристик. Следует сообщать обо всех анализах независимо от того, были ли они скорректированными или нет, и указывать, было ли проведение скорректированного анализа запланировано в протоколе и какие показатели были для этого выбраны. Авторы должны указать, планировалось ли проведение скорректированных анализов, включая выбор переменных для корректировки. В идеале в протоколе исследования должно быть указано, проводится ли корректировка на номинированные исходные переменные с помощью ковариационного анализа [72]. Корректировка на переменные, поскольку они существенно различаются на исходном уровне, скорее всего, приведет к искажению расчетного эффекта лечения [72]. Исследование показало, что неосознанные расхождения между протоколами и публикациями были обнаружены во всех 25 исследованиях, в которых были представлены результаты анализа подгрупп, и в 23 из 28 исследований,

в которых были представлены результаты скорректированных анализов [75].

Пункт 19. Указать все важные вредные или не преднамеренные последствия в каждой группе (для получения конкретных указаний см. CONSORT для вреда)

Для просмотра конкретных указаний обратитесь к CONSORT о побочных эффектах [76].

Пример: «Доли пациентов, у которых было отмечено возникновение любых неблагоприятных эффектов, в группе rBPI21 (рекомбинантный бактерицидный белок, повышающий проницаемость мембран) и группе плацебо были одинаковыми: 168 (88,4%) из 190 и 180 (88,7%) из 203 больных соответственно; при отдельной оценке неблагоприятных эффектов, классифицированных по 12 системам органов, доля побочных эффектов, относящихся к 11 системам органов, была несколько ниже в группе rBPI21. Доля больных, у которых развились тяжелые неблагоприятные эффекты, в группе rBPI21 была ниже, чем в группе плацебо: 53 (27,9%) из 190 и 74 (36,5%) из 203 больных соответственно. Только в 3 случаях возникновения тяжелых побочных эффектов сообщалось об их связи с применявшимся вмешательством; все они отмечались в группе плацебо» [79].

Объяснение. Для принятия рациональных и взвешенных решений читателям необходима информация как о вреде, так и о пользе вмешательств. Наличие и характер неблагоприятных эффектов могут существенно повлиять на то, будет ли то или иное вмешательство признано приемлемым и полезным. Не все зарегистрированные нежелательные явления, наблюдаемые в ходе исследования, обязательно являются следствием вмешательства; некоторые из них могут быть следствием заболевания, которое лечится. Рандомизированные исследования являются наилучшим подходом для получения данных о безопасности и эффективности, хотя они и не могут выявить редкие вредные последствия.

Многие статьи о РКИ содержат неадекватную информацию о побочных явлениях. Обзор 192 исследований лекарственных средств, опубликованных с 1967 по 1999 год, показал, что только 39% из них содержали адекватную информацию о клинических нежелательных явлениях и 29% — о токсичности, определяемой лабораторными методами [78]. Недавно сравнение данных о побочных явлениях, представленных в базу данных исследований Национального института рака, который спонсировал исследования, с информацией, опубликованной в журнальных статьях, показало, что в журнальных статьях занижены данные о побочных явлениях низкой степени тяжести. О событиях высокой степени тяжести (Common Toxicity Criteria, классы 3–5) в статьях сообщалось непоследовательно, а информация об отнесении их к исследуемым препаратам была неполной [80]. Более того, анализ исследований, опубликованных в шести общемедицинских журналах в 2006–2007 годах, показал, что, хотя в 89% из 133 статей упоминались нежелательные

явления, информация о тяжелых нежелательных явлениях и исключении пациентов из исследований в связи с нежелательными явлениями отсутствовала в 27% и 48% статей соответственно [81].

Было разработано расширенное заявление CONSORT, содержащее подробные рекомендации по представлению информации о вреде в рандомизированных исследованиях [84]. Рекомендации и примеры соответствующих статей находятся в свободном доступе на сайте CONSORT (www.consort-statement.org). Они дополняют Положение CONSORT 2010, и с ними следует ознакомиться, особенно если изучение вреда является ключевой задачей. Вкратце, если собирались данные о нежелательных явлениях, то они должны быть перечислены и определены, при необходимости со ссылкой на стандартизованные критерии. Должны быть описаны методы, использованные для сбора данных и отнесения событий. Для каждой группы исследования следует представить абсолютный риск каждого нежелательного явления, а также число участников, выбывших из исследования из-за причиненного вреда. Наконец, авторы должны представить сбалансированное обсуждение пользы и вреда [84].

Обсуждение

Пункт 20. Ограничения исследования, устранение источников потенциальной систематической ошибки, неточности и при необходимости множественности анализов

Пример: «Преобладание пациентов мужского пола (85%) является ограничением нашего исследования. Мы использовали стенты из чистого металла, поскольку стенты с лекарственным покрытием были доступны только в конце набора. Хотя последний фактор может восприниматься как ограничение, опубликованные данные не указывают на преимущества (как краткосрочные, так и долгосрочные) в отношении смертности и инфаркта миокарда у пациентов со стабильной ИБС, у которых устанавливают стенты с лекарственным покрытием, по сравнению с тем, кому устанавливают стенты из чистого металла» [82].

Объяснение. В дискуссионных разделах научных статей представлены заявления в поддержку выводов авторов [83], однако содержится недостаточное количество взвешенных аргументов, поддерживающих или опровергающих исследование и его результаты. Некоторые журналы пытались решить эту проблему, поощряя более структурированное обсуждение авторами своих результатов [85, 86]. Например, *Annals of Internal Medicine* рекомендует авторам структурировать раздел обсуждения, представляя: 1) краткий обзор основных результатов, 2) рассмотрение возможных механизмов и объяснений, 3) сравнение с соответствующими результатами других опубликованных исследований (по возможности включая систематический обзор, объединяющий результаты текущего исследования с результатами всех предыдущих соответствующих исследований), 4) ограничения настоящего исследования (и методы, используемые для минимизации и компенсации этих ограничений), 5) краткий

раздел, в котором резюмируются клинические и исследовательские последствия работы, в зависимости от обстоятельств [85]. Мы рекомендуем авторам следовать данным конструктивным предложениям, возможно, также используя подходящие подзаголовки в разделе «Обсуждения».

Хотя обсуждение ограничений часто опускается в статьях об исследованиях [87], выявление и обсуждение слабых сторон исследования имеют особое значение [88]. Например, хирургическая группа сообщила, что лапароскопическая холецистэктомия, технически сложная процедура, имеет значительно более низкую частоту осложнений, чем более традиционная открытая холецистэктомия для лечения острого холецистита [89]. Однако авторы не смогли обсудить очевидную предвзятость в своих результатах. Исследователи выполнили все лапароскопические холецистэктомии, тогда как 80% открытых холецистэктомий были выполнены стажерами.

Авторы также должны обсудить любую неточность результатов. Неточность может возникнуть в связи с несколькими аспектами исследования, включая измерение основного результата (см. пункт 6а) или диагностику (см. пункт 4а). Возможно, используемая шкала была валидирована на взрослой популяции, но использовалась в педиатрии, или оценщик не был обучен тому, как пользоваться прибором.

Следует всегда помнить о разнице между статистической значимостью и клинической значимостью. Авторы должны особенно избегать распространенной ошибки интерпретации незначительного результата как указания на эквивалентность вмешательств. Доверительный интервал (см. пункт 17а) дает ценную информацию о том, совместим ли результат исследования с клинически важным эффектом, независимо от значения P [90].

Авторам следует проявлять особую осторожность при оценке результатов исследований с множественными сравнениями. Такая множественность возникает из-за нескольких вмешательств, показателей результатов, моментов времени, анализа подгрупп и других факторов. В таких обстоятельствах некоторые статистически значимые результаты, вероятно, являются результатом одной лишь случайности.

Пункт 21. Обобщение (внешняя валидность, применимость) результатов исследования

Примеры: «Поскольку вмешательство проводилось для представителей обоих полов, всех возрастов, всех видов спорта и на разных спортивных уровнях, результаты показывают, что весь спектр спортсменов, от молодых выдающихся до спортсменов среднего и старшего возраста, получит пользу от выполнения разработанной программы тренировок по профилактике рецидивов растяжения связок голеностопного сустава. Включив спортсменов, не прошедших и прошедших лечение, мы охватили широкий спектр тяжести травм. Это говорит о том, что данная программа тренировок может быть реализована при лечении всех спортсменов. Кроме того, можно предположить, что

растяжения связок голеностопного сустава, не связанные со спортом, сравнимы с растяжениями в спорте, программа может принести пользу населению в целом» [91].

«Это повторяет и расширяет работу Кларка и его коллег и демонстрирует, что данная программа когнитивно-поведенческой профилактики может быть надежно и эффективно реализована в различных условиях врачами, не входящими в группу, изначально разрабатывавшую вмешательство. Величина эффекта согласовывалась с таковой в ранее опубликованных одноцентровых исследованиях по профилактике депрессии и была надежной в разных центрах в отношении как депрессивных расстройств, так и симптомов. В данном общем исследовании мы выбрали условия сравнения, имеющие отношение как к общественному здравоохранению, так и к обычному уходу. Выборка также включала преимущественно представителей рабочего и среднего классов с доступом к медицинскому страхованию. Учитывая доказательства того, что когнитивно-поведенческая терапия может быть более эффективной для подростков из семей с более высокими доходами, необходимо проверить эффекты этой профилактической программы на более разнообразных в экономическом и этническом отношении выборках [92]».

Объяснение. Внешняя валидность, также называемая обобщаемостью или применимостью, — это степень, с которой результаты исследования могут быть применимы в других обстоятельствах [93]. Внутренняя валидность (достоверность) — степень, в которой конструкция и проведение исследования исключают возможность смещения (ошибки); является предпосылкой для внешней достоверности, так как, когда результаты несовершенного исследования недостоверны, вопрос о его внешней валидности становится неактуальным. Не существует внешней валидности как таковой: этот термин имеет смысл только в отношении четко определенных условий, которые не были непосредственно изучены в ходе конкретного исследования. Могут ли быть применимы результаты данного исследования к данному больному или к группам больных определенного возраста, пола, с определенной стадией заболевания или с сопутствующими заболеваниями? Являются ли результаты данного исследования применимыми при использовании других препаратов той же фармакологической группы, других дозировок, схем и путей введения, а также сопутствующей терапии? Можно ли ожидать аналогичных результатов при использовании вмешательства на первичном, вторичном и третичном уровнях медицинской помощи? Как влияет вмешательство на другие клинические исходы, которые не были оценены в данном исследовании, и важна ли продолжительность наблюдения и лечения, особенно в отношении вреда? [94]

Внешняя валидность (достоверность) является предметом субъективной оценки и зависит от характеристик участников, включенных в исследование, места проведения исследования, применявшихся схем

лечения и оцениваемых клинических результатов [97, 98]. Поэтому крайне важно, чтобы в статье была представлена полная информация о критериях включения участников, клинической базе и месте проведения исследования (см. пункт 4b), о вмешательствах и схемах их применения (см. пункт 5), об определении клинических результатов (см. пункт 6), а также о начале и конце периодов включения и последующего наблюдения (см. пункт 14). Важно также сообщать о доле участников в контрольной группе, у которых развился тот или иной клинический исход (что позволяет оценить степень риска его возникновения в контрольной группе). Доля участников, которые отказываются входить в исследование, как указано на блок-схеме (см. пункт 13), имеет отношение к обобщению исследования, поскольку она может указывать на предпочтения или приемлемость вмешательства. Аналогичные соображения могут применяться к предпочтениям врачей [95, 96].

Некоторые из проблем важны, когда результаты исследования применяются к отдельному конкретному пациенту [99–101]. Эффективность лечения может быть разной у отдельных пациентов, участников одного или нескольких исследований, включенных в систематический обзор, однако такие различия, как правило, имеют скорее количественный, чем качественный характер.

Однако существуют важные исключения [101], терапия (особенно лекарственные препараты [102]), признанная полезной у узкого круга пациентов, как правило, имеет более широкое применение в реальной практике. Были предложены рамки для оценки внешней достоверности, включая качественные исследования, такие как интегральные «оценки процессов» [103] и контрольные списки [104]. При оценке соотношения пользы и риска определенного вмешательства у отдельного пациента или группы пациентов с характеристиками, отличающимися от характеристик участников исследования, полезно ориентироваться на рассчитанные авторами статьи показатели, учитывающие исходный риск и размер эффекта, как, например, число больных, которых необходимо лечить в течение определенного времени для получения еще одного благоприятного исхода, или число больных, которые должны получить лечение, чтобы развился еще один неблагоприятный исход [101, 105, 106]. Наконец, после получения ориентированных на пациента оценок потенциальной эффективности и безопасности от вмешательства врач должен учесть мнение и предпочтения пациента. Аналогичные соображения используются при оценке применимости результатов исследования в других условиях оказания медицинской помощи и при использовании других вмешательств.

Пункт 22. Интерпретация в соответствии с результатами, сопоставление преимуществ и недостатков, рассмотрение других значимых факторов

Пример: «Результаты исследований, опубликованных до 1990 года, показали, что профилактическая

иммунотерапия также снижает частоту внутрибольничных инфекций у новорожденных с очень низкой массой тела при рождении. Тем не менее в эти исследования было включено небольшое количество пациентов, были использованы различные дизайны, препараты и их дозы, а также включались разнообразные группы популяции. В данном крупном многоцентровом рандомизированном контролируемом исследовании повторное профилактическое внутривенное введение иммуноглобулина не сопровождалось статистически значимым снижением частоты внутрибольничных инфекций у недоношенных детей с массой тела от 501 до 1500 г при рождении» [107].

Объяснение. Читатели захотят узнать, как результаты настоящего исследования соотносятся с результатами других ранее опубликованных РКИ. Это может быть лучше всего достигнуто путем включения формального систематического обзора (метаанализа) в раздел «Результаты» или «Обсуждение» [108–110]. Такое обобщение доступно только при наличии опубликованных результатов предыдущих РКИ и часто практически невыполнимо. Систематический обзор помогает читателям оценить, аналогичны ли результаты РКИ других исследований по той же теме, а также дать ценную информацию о степени сходства между участниками разных исследований. В настоящее время такая информация в статьях о РКИ должным образом не рассматривается [107]. Методы Байеса могут быть использованы для статистического объединения результатов данных и предшествующих исследований [110].

Мы рекомендуем авторам обсуждать результаты РКИ, по крайней мере в рамках уже имеющихся доказательств. Это обсуждение должно быть систематическим и основываться на всестороннем поиске, а не ограничиваться цитированием исследований, подтверждающих результаты представляемого исследования [112].

Другая информация

Пункт 23. Регистрационный номер и название исследования в реестре

Пример: «Исследование зарегистрировано на ClinicalTrials.gov, номер NCT00244842» [113].

Объяснение. Последствия отсутствия публикации полных результатов исследований [114, 115], выборочной отчетности о результатах в ходе исследований и анализа каждого протокола, вместо анализа по назначенному лечению, были хорошо задокументированы [123, 124, 116]. Скрытая избыточная публикация клинических исследований также может вызвать проблемы, особенно для авторов систематических обзоров, когда результаты того же исследования, по видимому, случайно включены более чем один раз [117].

Чтобы свести к минимуму или избежать этих проблем, в течение последних 25 лет неоднократно осуществлялись призывы к регистрации клинических исследований в момент их начала для назначения уникальных идентификационных номеров исследований

и для записи другой базовой информации об исследовании, чтобы существенные детали были общедоступны [118–121]. Из-за возникших недавних серьезных проблем с хранением данных [122] были возобновлены попытки регистрации рандомизированных исследований. Действительно, Всемирная организация здравоохранения заявляет, что «регистрация всех интервенционных исследований является научной, этической и моральной ответственностью» (www.who.int/ictpr/en). Регистрируя рандомизированное исследование, авторы обычно сообщают о минимальном наборе информации и получают уникальный регистрационный номер исследования.

В сентябре 2004 года Международный комитет редакторов медицинских журналов (ICMJE — International Committee of Medical Journal Editors) изменил свою политику, заявив, что они будут рассматривать исследования для публикации только в том случае, если они были зарегистрированы до регистрации первого участника [125]. Это привело к резкому увеличению числа зарегистрированных исследований [126]. Международный комитет редакторов медицинских журналов дает рекомендации по допустимым реестрам (www.icmje.org/faq.pdf; http://www.icmje.org/faq_clinical.html).

В недавнем опросе 165 медицинских журналов с высоким показателем цитируемости статей (импакт-фактор), содержащих инструкции для авторов, 44 журнала конкретно заявили, что все недавние клинические исследования должны быть зарегистрированы в качестве требования к публикации в данном журнале [127].

Авторы должны указать название реестра и уникальный регистрационный номер исследования. Если авторы не зарегистрировали свое исследование, они должны четко указать это и назвать причину.

Пункт 24. Информация о получении доступа к полному протоколу исследования, при наличии

Пример: «Полную информацию о протоколе исследования можно найти в дополнительном приложении, доступном вместе с полным текстом этой статьи на сайте www.nejm.org» [128].

Объяснение. Протокол полного исследования (а не протокол конкретного этапа в рамках исследования) важен, потому что он заранее определяет методы рандомизированного исследования, такие как первичный исход (см. пункт 6a). Наличие протокола может помочь ограничить вероятность необъявленных изменений в методах исследования и выборочной отчетности о результатах (см. пункт 6b). Элементы, которые могут быть важны для включения в протокол рандомизированного исследования, описаны в других источниках [129].

Есть несколько вариантов, которые авторы могут рассмотреть, чтобы их протокол исследования был доступен заинтересованным читателям. Как описано в приведенном выше примере, журналы, сообщающие об основных результатах исследования, могут

разместить протокол исследования на своем веб-сайте. Доступ к результатам исследований и протоколу улучшается, когда журнал находится в открытом доступе. Некоторые журналы (например, *Trials*) публикуют протоколы исследований, и на такую публикацию можно ссылаться при сообщении об основных результатах исследования. Регистрация исследования (см. пункт 23) также обеспечит наличие многих подробностей протокола исследования, поскольку минимальные характеристики исследования, включенные в утвержденную базу данных регистрации исследования, включают несколько элементов протокола и результатов (www.who.int/ictpr/en). Исследователи могут также иметь возможность размещать протокол своего исследования на веб-сайте через работодателя. Какой бы механизм ни использовался, мы рекомендуем всем исследователям сделать свой протокол легкодоступным для заинтересованных читателей.

Пункт 25. Источники финансирования и другая поддержка (например, обеспечение лекарственными средствами), роль финансирующих организаций

Примеры: «Грантовая поддержка была получена для проведения вмешательства от Plan International и для исследования от Wellcome Trust и Объединенной программы Организации Объединенных Наций по ВИЧ/СПИДу (ЮНЭЙДС). Спонсоры не участвовали в разработке исследования, сборе и анализе данных, решении о публикации или подготовке манускрипта (рукописи)» [130].

«Это исследование финансировалось GlaxoSmithKline Pharmaceuticals. GlaxoSmithKline участвовала в разработке и проведении исследования и обеспечивала материально-техническую поддержку во время исследования. Сотрудники спонсора работали с исследователями над подготовкой плана статистического анализа, но сам анализ проводился Университетом штата Юта. Манускрипт (рукопись) была подготовлена доктором Шадди и членами руководящего комитета. GlaxoSmithKline было разрешено просматривать манускрипт (рукопись) и предлагать изменения, но окончательное решение по содержанию оставалось исключительно за авторами» [131].

Объяснение. Авторы должны сообщать об источниках финансирования исследования, так как это важная информация для читателей, оценивающих исследование. Анализ показал, что исследования, спонсируемые фармацевтической промышленностью, с большей вероятностью дадут результаты в пользу продукта, произведенного компанией, спонсирующей исследование, чем исследования, финансируемые из других источников [132–135]. Систематический обзор 30 исследований по вопросам финансирования показал, что исследования, финансируемые фармацевтической промышленностью, имеют в четыре раза больше шансов на достижение результатов в пользу спонсора, чем исследования, финансируемые из других источников (отношение шансов 4,05, 95% доверительный интервал от 2,98 до 5,51) [132]. Большая

часть публикаций об исследованиях в настоящее время не сообщает об источниках финансирования. Степень занижения трудно определить количественно. Обзор 370 исследований лекарственных средств показал, что 29% не сообщили об источниках финансирования [136]. В другом обзоре рандомизированных исследований, проиндексированных PubMed, опубликованном в декабре 2000 года, источник финансирования был указан для 66% из 519 исследований [138].

Уровень участия спонсора и его влияние на дизайн, проведение, анализ и отчетность исследования варьируются. Поэтому важно, чтобы авторы подробно описывали роль спонсоров. Если спонсор не имел такого участия, авторы должны заявить об этом. Точно так же авторы должны сообщать о любых других источниках поддержки, таких как поставка и подготовка лекарственных препаратов или оборудования, или при анализе данных и написании рукописи [137].

Отчетность рандомизированных контролируемых исследований, в которых не было двухгруппового параллельного дизайна

Основное внимание в рекомендациях CONSORT уделяется РКИ с параллельным дизайном и двумя группами лечения. Хотя большинство исследований построены именно таким образом, также существует значимое меньшинство: отличный дизайн имели 45% (233/519) РКИ, опубликованных в декабре 2000 года [138], и 39% (242/616) в декабре 2006 года [17].

Большая часть заявления CONSORT в равной степени применима ко всем дизайнам исследований, но есть несколько дополнительных вопросов, которые необходимо решить для каждого дизайна. Перед публикацией пересмотренного заявления CONSORT в 2001 году группа CONSORT решила разработать дополнения к основному заявлению, относящиеся к конкретным дизайнам исследований. Были опубликованы расширения, касающиеся отчетов о кластерных рандомизированных исследованиях [139] и исследованиях не меньшей эффективности и эквивалентности [140]. Нехватка ресурсов означает, что другие запланированные расширения не были завершены; они будут охватывать исследования со следующими дизайнами: параллельные с несколькими группами, факторные, перекрестные, внутриличностные.

Авторы, сообщающие об исследованиях с кластерным дизайном или с использованием схемы не меньшей эффективности или эквивалентности, должны ознакомиться с рекомендациями CONSORT в дополнение к рекомендациям, содержащимся в этом документе. Здесь мы делаем несколько промежуточных замечаний о других проектах. В каждом случае дизайн исследования должен быть четко изложен как в основном тексте, так и в аннотации статьи.

Многогрупповые (> 2 групп) исследования в параллельных группах требуют минимальной переработки стандартного руководства CONSORT. Блок-схему можно легко расширить. Основные отличия от исследований с двумя группами касаются уточнения того, как гипотезы исследования соотносятся

с несколькими группами, и последующих методов анализа и интерпретации данных. Для факторных исследований обычно необходимо учитывать возможность взаимодействия между вмешательствами. В дополнение к общему сравнению участников, получавших или не получавших каждое исследуемое вмешательство, исследователи должны также рассмотреть возможность сообщения результатов для каждой комбинации лечения [141].

В перекрестных исследованиях каждый участник получает два (или более) типа лечения в случайном порядке. Основные дополнительные проблемы, которые необходимо решить, связаны с парным характером данных, которые влияют на дизайн и анализ [142]. Аналогичные проблемы затрагивают сравнения внутри человека, в которых участники получают два лечения одновременно (часто на парные органы). Кроме того, из-за риска временных или системных эффектов переноса, соответственно, в обоих случаях выбор дизайна требует обоснования.

Группа CONSORT намерена опубликовать расширение к заявлению CONSORT для охвата всех указанных дизайнов. Дополнительно мы опубликуем обновления существующих руководств по кластерным рандомизированным исследованиям и исследованиям не меньшей эффективности и эквивалентности, чтобы учесть это крупное обновление общего руководства CONSORT.

ОБСУЖДЕНИЕ

Оценка медицинских вмешательств может вводить в заблуждение, если исследователи не обеспечивают объективность сравнений. Случайное распределение по группам исследования остается единственным методом, исключающим предвзятость отбора и сбивающие факторы. Нерандомизированные исследования, как правило, дают больший расчетный эффект лечения, чем рандомизированные [143, 144].

Однако предвзятость ставит под угрозу даже рандомизированные контролируемые исследования, если ученые проводят их ненадлежащим образом [145]. Недавний систематический обзор, объединивший результаты нескольких методологических исследований, показал, что исследования, в которых использовалось неадекватное или нечеткое сокрытие распределения, дали на 31% большую оценку эффекта, чем те, в которых использовалось адекватное сокрытие, а исследования, в которых не проводилось ослепление, дали на 25% большую оценку [149]. Как и следовало ожидать, между этими двумя показателями существует сильная связь.

Разработка и проведение РКИ требуют как методологических, так и клинических знаний, кропотливой работы [146] и высокой готовности к непредвиденным трудностям. Отчеты о РКИ должны быть написаны с таким же пристальным вниманием к снижению предвзятости. Читатели не должны строить догадки; используемые методы должны быть полными и прозрачными, чтобы читатели могли легко отличить

исследования с объективными результатами от сомнительных. Разумная наука включает в себя адекватную отчетность, а проведение этических испытаний опирается на фундамент разумной науки [147].

Мы надеемся, что данное обновление пояснительной статьи CONSORT поможет авторам в использовании версии CONSORT 2010 г. и в общих чертах объяснит важность адекватного освещения исследований. Положение CONSORT может помочь исследователям, разрабатывающим испытания в будущем [148], и может служить руководством для рецензентов и редакторов при оценке рукописей. Действительно, мы призываем рецензентов и редакторов использовать контрольный список CONSORT для оценки того, насколько полно авторы представили информацию по этим пунктам. Такая оценка, вероятно, повысит ясность и прозрачность публикуемых исследований. Поскольку CONSORT является развивающимся документом, он требует динамичного процесса постоянной оценки, доработки и при необходимости внесения изменений, именно поэтому мы подготовили данное обновление контрольного перечня и пояснительной статьи. По мере накопления новых данных и критических замечаний мы будем оценивать необходимость будущих обновлений.

Первая версия заявления CONSORT, выпущенная в 1996 году, по-видимому, привела к улучшению качества отчетности о РКИ в журналах, которые ее приняли [150–154]. Другие группы используют шаблон CONSORT для улучшения отчетности по другим видам исследований, таким как диагностические тесты [155] и обсервационные исследования [156].

Сайт CONSORT (www.consort-statement.org) был создан с целью предоставления образовательных материалов и базы данных, содержащей материалы, относящиеся к представлению отчетов о РКИ. На сайте представлено множество примеров из реальных исследований, в том числе все примеры, включенные в данную статью. Мы будем продолжать пополнять базу данных хорошими и плохими примерами отчетности и приглашаем читателей вносить дополнительные предложения, связываясь с нами через сайт. Группа CONSORT будет продолжать изучать литературу с целью поиска соответствующих статей, в которых рассматриваются вопросы, связанные с представлением результатов РКИ, и мы приглашаем авторов таких статей сообщать нам о них. Эта информация будет доступна на регулярно обновляемом сайте CONSORT.

Официальную поддержку CONSORT выразили более 400 ведущих общих и специализированных журналов и редакционных групп по биомедицине, включая Международный комитет редакторов медицинских журналов, Всемирную ассоциацию редакторов медицинских журналов и Совет научных редакторов. Мы приглашаем другие журналы, заинтересованные в качественном освещении клинических исследований, поддержать заявление CONSORT и связаться с нами через наш сайт, чтобы сообщить о своей поддержке. Конечными бенефициарами этих

коллективных усилий должны стать люди, которые по тем или иным причинам нуждаются во вмешательстве со стороны медицинского сообщества.

СПИСОК ЛИТЕРАТУРЫ

- [1] Tonino PA, De Bruyne B, Pijls NH, Siebert U, Ikeno F, Veer M, et al. Fractional flow reserve versus angiography for guiding percutaneous coronary intervention. *N Engl J Med* 2009;360:213-24.
- [2] Grant AM, Wileman SM, Ramsay CR, Mowat NA, Krukowski ZH, Heading RC, et al. Minimal access surgery compared with medical management for chronic gastro-oesophageal reflux disease: UK collaborative randomised trial. *BMJ* 2008;337:a2664.
- [3] Egger M, Juni P, Bartlett C. Value of flow diagrams in reports of randomized controlled trials. *JAMA* 2001;285:1996-9.
- [4] Altman DG. Randomisation. *BMJ* 1991;302:1481-2.
- [5] Sackett DL, Gent M. Controversy in counting and attributing events in clinical trials. *N Engl J Med* 1979;301:1410-2.
- [6] May GS, DeMets DL, Friedman LM, Furberg C, Passamani E. The randomized clinical trial: bias in analysis. *Circulation* 1981;64:669-73.
- [7] Altman DG, Cuzick J, Peto J. More on zidovudine in asymptomatic HIV infection. *N Engl J Med* 1994;330:1758-9.
- [8] Meinert CL. Beyond CONSORT: need for improved reporting standards for clinical trials. *Consolidated Standards of Reporting Trials*. *JAMA* 1998;279:1487-9.
- [9] 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.
- [10] 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.
- [11] Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;357:1191-4.
- [12] 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.
- [13] van Loon AJ, Mantingh A, Serlier EK, Kroon G, Mooyaart EL, Huisjes HJ. Randomised controlled trial of magnetic-resonance pelvimetry in breech presentation at term. *Lancet* 1997;350:1799-804.
- [14] Brown MJ, Palmer CR, Castaigne A, de Leeuw PW, Mancia G, Rosenthal T, et al. Morbidity and mortality in patients randomised to double-blind treatment with a long-acting calcium-channel blocker or diuretic in the International Nifedipine GITS study: Intervention as a Goal in Hypertension Treatment (INSIGHT). *Lancet* 2000;356:366-72.
- [15] LaCroix AZ, Ott SM, Ichikawa L, Scholes D, Barlow WE. Low-dose hydrochlorothiazide and preservation of bone mineral density in older adults. A randomized, double-blind, placebo-controlled trial. *Ann Intern Med* 2000;133:516-26.
- [16] Shuster JJ. Median follow-up in clinical trials. *J Clin Oncol* 1991;9:191-2.
- [17] Altman DG, de Stavola BL, Love SB, Stepniowska KA. Review of survival analyses published in cancer journals. *Br J Cancer* 1995;72:511-8.
- [18] Auvert B, Taljaard D, Lagarde E, Sobngwi-Tambekou J, Sitta R, Puren A. Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 Trial. *PLoS Med* 2005;2:e298.
- [19] Diggle L, Deeks J. Effect of needle length on incidence of local reactions to routine immunisation in infants aged 4 months: randomised controlled trial. *BMJ* 2000;321:931-3.
- [20] Pocock S, White I. Trials stopped early: too good to be true? *Lancet* 1999;353:943-4.
- [21] 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.
- [22] Hughes MD, Pocock SJ. Stopping rules and estimation problems in clinical trials. *Stat Med* 1988;7:1231-42.

- [23] Kiri A, Tonascia S, Meinert CL. Treatment effects monitoring committees and early stopping in large clinical trials. *Clin Trials* 2004;1: 40-7.
- [24] Psaty BM, Rennie D. Stopping medical research to save money: a broken pact with researchers and patients. *JAMA* 2003;289: 2128-31.
- [25] Altman DG, Dore CJ. Randomisation and baseline comparisons in clinical trials. *Lancet* 1990;335:149-53.
- [26] 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.
- [27] Lang TA, Secic M. How to report statistics in medicine. Annotated guidelines for authors, editors, and reviewers. ACP 1997.
- [28] 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.
- [29] Senn S. Base logic: tests of baseline balance in randomized clinical trials. *Clin Res Regulatory Affairs* 1995;12:171-82.
- [30] Altman DG. Comparability of randomised groups. *Statistician* 1985;34:125-36.
- [31] Yusuf S, Teo K, Anderson C, Pogue J, Dyal L, Copland I, et al. Effects of the angiotensin-receptor blocker telmisartan on cardiovascular events in high-risk patients intolerant to angiotensin-converting enzyme inhibitors: a randomised controlled trial. *Lancet* 2008;372:1174-83.
- [32] Heit JA, Elliott CG, Trowbridge AA, Morrey BF, Gent M, Hirsh J. Ardeparin sodium for extended out-of-hospital prophylaxis against venous thromboembolism after total hip or knee replacement. A randomized, double-blind, placebo-controlled trial. *Ann Intern Med* 2000;132:853-61.
- [33] Haderslev KV, Tjellesen L, Sorensen HA, Staun M. Alendronate increases lumbar spine bone mineral density in patients with Crohn's disease. *Gastroenterology* 2000;119:639-46.
- Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 1999;319: 670-4.
- [34] Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 1999;319: 670-4.
- [35] Fields WS, Maslenikov V, Meyer JS, Hass WK, Remington RD, Macdonald M. Joint study of extra-cranial arterial occlusion. V. Progress report of prognosis following surgery or nonsurgical treatment for transient cerebral ischemic attacks and cervical carotid artery lesions. *JAMA* 1970;211:1993-2005.
- [36] Lee YJ, Ellenberg JH, Hirtz DG, Nelson KB. Analysis of clinical trials by treatment actually received: is it really an option? *Stat Med* 1991;10:1595-605.
- [37] Lewis JA, Machin D. Intention to treat who should use ITT? *Br J Cancer* 1993;68:647-50.
- [38] Lachin JL. Statistical considerations in the intent-to-treat principle. *Control Clin Trials* 2000;21:526.
- [39] Sheiner LB, Rubin DB. Intention-to-treat analysis and the goals of clinical trials. *Clin Pharmacol Ther* 1995;57:6-15.
- [40] Nagelkerke N, Fidler V, Bernsen R, Borgdorff M. Estimating treatment effects in randomized clinical trials in the presence of noncompliance. *Stat Med* 2000;19:1849-64.
- [41] Melander H, Hlqvist-Rastad J, Meijer G, Beermann B. Evidence based medicine-selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* 2003;326:1171-3.
- [42] Gravel J, Opatrný L, Shapiro S. The intention-to-treat approach in randomized controlled trials: are authors saying what they do and doing what they say? *Clin Trials* 2007;4:350-6.
- [43] Kruse RL, Alper BS, Reust C, Stevermer JJ, Shannon S, Williams RH. Intention-to-treat analysis: who is in? Who is out? *J Fam Pract* 2002;51:969-71.
- [44] Herman A, Botser IB, Tenenbaum S, Chechick A. Intention-to-treat analysis and accounting for missing data in orthopaedic randomized clinical trials. *J Bone Joint Surg Am* 2009;91:2137-43.
- [45] 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.
- [46] Ruiz-Canela M, Martinez-Gonzalez MA, de Irala-Estevez J. Intention to treat analysis is related to methodological quality. *BMJ* 2000;320:1007-8.
- [47] Altman DG. Missing outcomes in randomised trials: addressing the dilemma. *Open Med* 2009;3:e21-3.
- [48] Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clin Trials* 2004;1: 368-76.
- [49] Streiner DL. Missing data and the trouble with LOCF. *Evid Based Ment Health* 2008;11:3-5.
- [50] Altman DG. Confidence intervals in practice. In: Altman DG, Machin D, Bryant TN, Gardner MJ, editors. *Statistics with confidence*. 2nd ed. BMJ Books; 2000. p. 6-14.
- [51] Lang TA, Secic M. How to report statistics in medicine. Annotated guidelines for authors, editors, and reviewers. ACP 1997.
- [52] Lang TA, Secic M. How to report statistics in medicine. Annotated guidelines for authors, editors, and reviewers. ACP 1997.
- [53] Uniform requirements for manuscripts submitted to biomedical journals. International Committee of Medical Journal Editors. *Ann Intern Med* 1997;126:36-47.
- [54] Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *BMJ* 1986;292:746-50.
- [55] Bailar JC III, Mosteller F. Guidelines for statistical reporting in articles for medical journals. Amplifications and explanations. *Ann Intern Med* 1988;108:266-73.
- [56] 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.
- [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] Bland JM. Quoting intermediate analyses can only mislead. *BMJ* 1997;314:1907-8.
- [59] Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *BMJ* 1995;310:452-4.
- [60] Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. *BMJ* 1999;319: 1492-5.
- [61] 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.
- [62] van Linschoten R, van Middelkoop M, Berger MY, Heintjes EM, Verhaar JA, Willemsen SP, et al. Supervised exercise therapy versus usual care for patellofemoral pain syndrome: an open label randomised controlled trial. *BMJ* 2009;339:b4074.
- [63] The OSIRIS Collaborative Group. Early versus delayed neonatal administration of a synthetic surfactant — the judgment of OSIRIS (open study of infants at high risk of or with respiratory insufficiency — the role of surfactant). *Lancet* 1992;340:1363-9.
- [64] Sorensen L, Gyrd-Hansen D, Kristiansen IS, Nexoe J, Nielsen JB. Laypersons' understanding of relative risk reductions: randomised cross-sectional study. *BMC Med Inform Decis Mak* 2008;8:31.
- [65] Bobbio M, Demichelis B, Giustetto G. Completeness of reporting trial results: effect on physicians' willingness to prescribe. *Lancet* 1994;343:1209-11.
- [66] Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? *Ann Intern Med* 1992;117:916-21.
- [67] 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.
- [68] Tukey JW. Some thoughts on clinical trials, especially problems of multiplicity. *Science* 1977;198:679-84.

- [69] Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA* 1991;266:93-8.
- [70] 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.
- [71] Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med* 1992;116:78-84.
- [72] Altman DG. Adjustment for covariate imbalance. In: Armitage P, Colton T, editors. *Encyclopedia of biostatistics*. John Wiley; 1998. p. 1000-5.
- [73] Hahn S, Williamson PR, Hutton JL, Garner P, Flynn EV. Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. *Stat Med* 2000;19:3325-36.
- [74] Bhandari M, Devereaux PJ, Li P, Mah D, Lim K, Schunemann HJ, et al. Misuse of baseline comparison tests and subgroup analyses in surgical trials. *Clin Orthop Relat Res* 2006;447:247-51.
- [75] 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.
- [76] 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.
- [77] Levin M, Quint PA, Goldstein B, Barton P, Bradley JS, Shemie SD, et al. Recombinant bactericidal/permeability-increasing protein (rBPI21) as adjunctive treatment for children with severe meningococcal sepsis: a randomised trial. *rBPI21 Meningococcal Sepsis Study Group. Lancet* 2000;356:961-7.
- [78] Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *JAMA* 2001;285:437-43.
- [79] 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.
- [80] Scharf O, Colevas AD. Adverse event reporting in publications compared with sponsor database for cancer clinical trials. *J Clin Oncol* 2006;24:3933-8.
- [81] Pitrou I, Boutron I, Ahmad N, Ravaud P. Reporting of safety results in published reports of randomized controlled trials. *Arch Intern Med* 2009;169:1756-61.
- [82] Boden WE, O'Rourke RA, Teo KK, Hartigan PM, Maron DJ, Kostuk WJ, et al. Optimal medical therapy with or without PCI for stable coronary disease. *N Engl J Med* 2007;356:1503-16.
- [83] Horton R. The rhetoric of research. *BMJ* 1995;310:985-7.
- [84] 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.
- [85] *Annals of Internal Medicine*. Information for authors. Available at www.annals.org (accessed 15 Jan 2008).
- [86] Docherty M, Smith R. The case for structuring the discussion of scientific papers. *BMJ* 1999;318:1224-5.
- [87] Purcell GP, Donovan SL, Davidoff F. Changes to manuscripts during the editorial process: characterizing the evolution of a clinical paper. *JAMA* 1998;280:227-8.
- [88] Ioannidis JP. Limitations are not properly acknowledged in the scientific literature. *J Clin Epidemiol* 2007;60:324-9.
- [89] Kiviluoto T, Siren J, Luukkonen P, Kivilaakso E. Randomised trial of laparoscopic versus open cholecystectomy for acute and gangrenous cholecystitis. *Lancet* 1998;351:321-5.
- [90] Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995;311:485.
- [91] Hupperets MD, Verhagen EA, van Mechelen W. Effect of unsupervised home based proprioceptive training on recurrences of ankle sprain: randomised controlled trial. *BMJ* 2009;339. b2684.
- [92] Garber J, Clarke GN, Weersing VR, Beardslee WR, Brent DA, Gladstone TR, et al. Prevention of depression in at-risk adolescents: a randomized controlled trial. *JAMA* 2009;301:2215-24.
- [93] Campbell D. Factors relevant to the validity of experiments in social settings. *Psychol Bull* 1957;54:297-312.
- [94] Rothwell PM. Factors that can affect the external validity of randomised controlled trials. *PLoS Clin Trials* 2006;1:e9.
- [95] King M, Nazareth I, Lampe F, Bower P, Chandler M, Morou M, et al. Conceptual framework and systematic review of the effects of participants' and professionals' preferences in randomised controlled trials. *Health Technol Assess* 2005;9:1-iv.
- [96] Djulbegovic B, Lacey M, Cantor A, Fields KK, Bennett CL, Adams JR, et al. The uncertainty principle and industry-sponsored research. *Lancet* 2000;356:635-8.
- [97] Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ* 2001;323:42-6.
- [98] 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.
- [99] Dans AL, Dans LF, Guyatt GH, Richardson S. Users' guides to the medical literature: XIV. How to decide on the applicability of clinical trial results to your patient. *Evidence-Based Medicine Working Group. JAMA* 1998;279:545-9.
- [100] Smith GD, Egger M. Who benefits from medical interventions? *BMJ* 1994;308:72-4.
- [101] McAlister FA. Applying the results of systematic reviews at the bedside. In: Egger M, Davey Smith G, Altman DG, editors. *Systematic reviews in health care: meta-analysis in context*. BMJ Books; 2001.
- [102] Bartlett C, Doyal L, Ebrahim S, Davey P, Bachmann M, Egger M, et al. The causes and effects of socio-demographic exclusions from clinical trials. *Health Technol Assess* 2005;9:iii-x,1.
- [103] Bonell C, Oakley A, Hargreaves J, Strange V, Rees R. Assessment of generalisability in trials of health interventions: suggested framework and systematic review. *BMJ* 2006;333:346-9.
- [104] Bornhoft G, Moxion-Bergemann S, Wolf U, Kienle GS, Michalsen A, Vollmar HC, et al. Checklist for the qualitative evaluation of clinical studies with particular focus on external validity and model validity. *BMC Med Res Methodol* 2006;6:56.
- [105] Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med* 1988;318:1728-33.
- [106] Altman DG. Confidence intervals for the number needed to treat. *BMJ* 1998;317:1309-12.
- [107] Fanaroff AA, Korones SB, Wright LL, Wright EC, Poland RL, Bauer CB, et al. A controlled trial of intravenous immune globulin to reduce nosocomial infections in very-low-birth-weight infants. National Institute of Child Health and Human Development Neonatal Research Network. *N Engl J Med* 1994;330:1107-15.
- [108] Randomised trial of intravenous atenolol among 16 027 cases of suspected acute myocardial infarction: ISIS-1. First International Study of Infarct Survival Collaborative Group. *Lancet* 1986;2:57-66.
- [109] Gotzsche PC, Gjørup I, Bonnen H, Brahe NE, Becker U, Burcharth F. Somatostatin v placebo in bleeding oesophageal varices: randomised trial and meta-analysis. *BMJ* 1995;310:1495-8.
- [110] Clarke M, Hopewell S, Chalmers I. Reports of clinical trials should begin and end with up-to-date systematic reviews of other relevant evidence: a status report. *J R Soc Med* 2007;100:187-90.
- [111] Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. *Ann Intern Med* 1999;130:995-1004.
- [112] Gotzsche PC. Reference bias in reports of drug trials. *BMJ* 1987;295:654-6.
- [113] Papp K, Bissonnette R, Rosoph L, Wasel N, Lynde CW, Searles G, et al. Efficacy of ISA247 in plaque psoriasis: a randomised, multicentre, double-blind, placebo-controlled phase III study. *Lancet* 2008;371:1337-42.
- [114] Dickersin K. How important is publication bias? A synthesis of available data. *AIDS Educ Prev* 1997;9:15-21.
- [115] Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technol Assess* 2000;4:1-115.
- [116] Williamson PR, Gamble C. Identification and impact of outcome selection bias in meta-analysis. *Stat Med* 2005;24:1547-61.

- [117] Tramer MR, Reynolds DJ, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. *BMJ* 1997;315:635-40.
- [118] Times RJ. Publication bias: the case for an international registry of clinical trials. *J Clin Oncol* 1986;4:1529-41.
- [119] Chalmers I. From optimism to disillusion about commitment to transparency in the medico-industrial complex. *J R Soc Med* 2006;99:337-41.
- [120] Tonks A. A clinical trials register for Europe. *BMJ* 2002;325:1314-5.
- [121] Dickensin K, Rennie D. Registering clinical trials. *JAMA* 2003;290: 516-23.
- [122] Whittington CJ, Kendall T, Fonagy P, Cottrell D, Cotgrove A, Boddington E. Selective serotonin reuptake inhibitors in childhood depression: systematic review of published versus unpublished data. *Lancet* 2004;363:1341-5.
- [123] 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.
- [124] 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.
- [125] De Angelis CD, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al. Is this clinical trial fully registered? A statement from the International Committee of Medical Journal Editors. *Lancet* 2005;365:1827-9.
- [126] Zarin DA, Ide NC, Tse T, Harlan WR, West JC, Lindberg DA. Issues in the registration of clinical trials. *JAMA* 2007;297:2112-20.
- [127] Hopewell S, Altman DG, Moher D, Schulz KF. Endorsement of the CONSORT Statement by high impact factor medical journals: a survey of journal editors and journal 'instructions to authors'. *Trials* 2008;9:20.
- [128] Russell JA, Walley KR, Singer J, Gordon AC, Hebert PC, Cooper DJ, et al. Vasopressin versus norepinephrine infusion in patients with septic shock. *N Engl J Med* 2008;358:877-87.
- [129] Chan AW, Tetzlaff J, Altman D, Gotzsche PC, Hrobjartsson A, Krleza-Jeric K, et al. The SPIRIT initiative: defining standard protocol items for randomised trials. Oral presentation at the 16th Cochrane Colloquium: Evidence in the era of globalisation; 2008 Oct 3-7; Freiburg, Germany [abstract]. *Zeitschrift fur Evidenz, Fortbildung und Qualitat im Gesundheitswesen* 2008;102:27.
- [130] Gregson S, Adamson S, Papaya S, Mundondo J, Nyamukapa CA, Mason PR, et al. Impact and process evaluation of integrated community and clinic-based HIV-1 control: a cluster-randomised trial in eastern Zimbabwe. *PLoS Med* 2007;4:e102.
- [131] Shaddy RE, Boucek MM, Hsu DT, Boucek RJ, Canter CE, Mahony L, et al. Carvedilol for children and adolescents with heart failure: a randomized controlled trial. *JAMA* 2007;298:1171-9.
- [132] Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003;326:1167-70.
- [133] Kjaergard LL, Als-Nielsen B. Association between competing interests and authors' conclusions: epidemiological study of randomised clinical trials published in the *BMJ*. *BMJ* 2002;325:249.
- [134] Bero L, Oostvogel F, Bacchetti P, Lee K. Factors associated with findings of published trials of drug-drug comparisons: why some statins appear more efficacious than others. *PLoS Med* 2007;4: e184.
- [135] Sismondo S. Pharmaceutical company funding and its consequences: a qualitative systematic review. *Contemp Clin Trials* 2008;29:109-13.
- [136] Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA* 2003;290:921-8.
- [137] Ross JS, Hill KP, Egilman DS, Krumholz HM. Guest authorship and ghostwriting in publications related to rofecoxib: a case study of industry documents from rofecoxib litigation. *JAMA* 2008;299: 1800-12.
- [138] Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet* 2005;365:1159-62.
- [139] Campbell MK, Elbourne DR, Altman DG. CONSORT statement: extension to cluster randomised trials. *BMJ* 2004;328:702-8.
- [140] 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.
- [141] McAlister FA, Straus SE, Sackett DL, Altman DG. Analysis and reporting of factorial trials: a systematic review. *JAMA* 2003;289: 2545-53.
- [142] Senn S. Crossover trials in clinical research. 2nd ed. Wiley; 2002.
- [143] Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovich C, Song F, et al. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;7. iii-173.
- [144] Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev* 2007;. MR000012.
- [145] Collins R, MacMahon S. Reliable assessment of the effects of treatment on mortality and major morbidity, I: clinical trials. *Lancet* 2001;357:373-80.
- [146] Schulz KF. Randomised trials, human nature, and reporting guidelines. *Lancet* 1996;348:596-8.
- [147] Murray GD. Promoting good research practice. *Stat Methods Med Res* 2000;9:17-24.
- [148] Narahari SR, Ryan TJ, Aggithaya MG, Bose KS, Prasanna KS. Evidence-based approaches for the Ayurvedic traditional herbal formulations: toward an Ayurvedic CONSORT model. *J Altern Complement Med* 2008;14:769-76.
- [149] 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.
- [150] 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.
- [151] Egger M, Juni P, Bartlett C. Value of flow diagrams in reports of randomized controlled trials. *JAMA* 2001;285:1996e9.
- [152] 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.
- [153] 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.
- [154] Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;357:1191-4.
- [155] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med* 2003;138:40-4.
- [156] von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007;147: 573-7.
- [157] Molnar FJ, Hutton B, Fergusson D. Does analysis using «last observation carried forward» introduce bias in dementia research? *CMAJ* 2008;179:751-3.
- [158] Ware JH. Interpreting incomplete data in studies of diet and weight loss. *N Engl J Med* 2003;348:2136-7.
- [159] Abraha I, Montedori A, Romagnoli C. Modified intention to treat: frequency, definition and implication for clinical trials [abstract]. Sao Paulo, Brazil: XV Cochrane Colloquium; 2007:86-7.