

<https://doi.org/10.15690/vsp.v21i3.2426>

J.P. Vandembroucke¹, E. von Elm^{2, 3}, D.G. Altman⁴, P.C. Gøtzsche⁵, C.D. Mulrow⁶, S.J. Pocock⁷, C. Poole⁸, J.J. Schlesselman^{9, 10}, M. Egger^{2, 11} for the STROBE Initiative

¹ Leiden University Medical Center, Лейден, Нидерланды

² University of Bern, Берн, Швейцария

³ University Medical Centre, Фрайбург, Германия

⁴ Cancer Research UK/NHS Centre for Statistics in Medicine, Оксфорд, Великобритания

⁵ Nordic Cochrane Centre, Rigshospitalet, Копенгаген, Дания

⁶ University of Texas Health Science Center, Сан-Антонио, США

⁷ London School of Hygiene and Tropical Medicine, Лондон, Великобритания

⁸ University of North Carolina School of Public Health, Чапел-Хилл, США

⁹ University of Pittsburgh Graduate School of Public Health, Питтсбург, США

¹⁰ University of Pittsburgh Cancer Institute, Питтсбург, США

¹¹ Department of Social Medicine, University of Bristol, Бристоль, Великобритания

Повышение качества отчетов о наблюдательных исследованиях в эпидемиологии (STROBE): разъяснения и уточнения

Контактная информация:

Matthias Egger, Институт социальной и профилактической медицины (ISPM) Бернского университета, кафедра социальной медицины Бристольского университета, e-mail: strobe@ispm.unibe.ch

Статья поступила: 25.05.2022, принята к печати: 14.06.2022

Большинство медицинских исследований являются наблюдательными (observational). Сообщения о таких исследованиях часто невысокого качества, что затрудняет оценку сильных и слабых сторон работы, а также обобщаемости (generalisability) ее результатов. Принимая во внимание эмпирические свидетельства и теоретические соображения, группа методологов, исследователей и научных редакторов разработала рекомендации «Повышение качества отчетов о наблюдательных исследованиях в эпидемиологии (STROBE): разъяснения и уточнения». Рекомендации STROBE содержат 22 пункта, связанных с оформлением следующих разделов научных статей: название, аннотация, введение, методы, результаты и их обсуждение, при этом 18 пунктов являются общими для когортных исследований (cohort studies), исследований «случай-контроль» (case-control studies) и одномоментных исследований (cross-sectional studies); 4 пункта специфичны для каждого из указанных дизайнов исследований (study designs). STROBE — руководство для авторов, необходимое для повышения качества отчетов о наблюдательных исследованиях, облегчающее критическую оценку исследования и его интерпретацию рецензентами, редакторами журналов и читателями. Цель этой разъясняющей и уточняющей статьи — способствовать более широкому применению, пониманию и распространению стандартов STROBE. В ней даются разъяснение смысла и обоснование применения каждого пункта руководства (checklist). По каждому пункту приводятся один или несколько опубликованных примеров правильного представления исследований и, при возможности, библиографические ссылки на подходящие эмпирические исследования и методологическую литературу. Представлены примеры потоковых диаграмм (flow diagrams) для описания последовательности исследования. Рекомендации STROBE, настоящая статья и соответствующий веб-сайт (<http://www.strobe-statement.org/>) должны стать полезным источником для повышения качества отчетов о результатах наблюдательных исследований.

Настоящая статья является переводом оригинальной публикации под редакцией д.м.н. Р.Т. Сайгитова. Перевод впервые опубликован в *Digital Diagnostics*. doi: 10.17816/DD70821. Публикуется с незначительными изменениями, связанными с литературным редактированием текста перевода.

Ключевые слова: STROBE, наблюдательные исследования, когортные исследования, одномоментные исследования, случай-контроль, дизайн исследования

Для цитирования (перевод): Jan P. Vandembroucke, Erik von Elm, Douglas G. Altman, Peter C. Gøtzsche, Cynthia D. Mulrow, Stuart J. Pocock, Charles Poole, James J. Schlesselman, Matthias Egger. Повышение качества отчетов о наблюдательных исследованиях в эпидемиологии (STROBE): разъяснения и уточнения. *Вопросы современной педиатрии*. 2022;21(3):173–208. doi: <https://doi.org/10.15690/vsp.v21i3.2426>

Для цитирования (оригинальная статья): Vandembroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, et al. (2007) Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *PLOS Medicine* 4(10): e297. <https://doi.org/10.1371/journal.pmed.0040297>

СПИСОК СОКРАЩЕНИЙ

ДИ — доверительный интервал

ИОР — избыточный относительный риск (вследствие взаимодействия)

ОР — относительный риск

STROBE — повышение качества отчетов о наблюдательных исследованиях в эпидемиологии

ВВЕДЕНИЕ

Осмысленная практика оказания медицинской помощи требует знания этиологии и патогенеза, диагностики, прогнозирования и терапии заболеваний. Рандомизированные испытания дают ценные сведения о методах лечения и иных вмешательствах. Однако существенная часть клинических данных, а также информация в области общественного здравоохранения основываются на наблюдательных исследованиях [1]. Примерно 9 из 10 научных работ, опубликованных в медицинских журналах, описывают наблюдательные исследования [2, 3].

Рекомендации STROBE

Отчеты о наблюдательных исследованиях часто бывают недостаточно подробными и четкими, что не позволяет оценить сильные и слабые стороны работы [4, 5]. Для повышения качества таких отчетов мы разработали правила оформления разделов статей — названия, аннотации, введения, методов, результатов и их обсуждения (табл. 1). Эти правила опубликованы в нескольких журналах [6]. Наша цель — обеспечить четкое представление того, что было запланировано, сделано и обнаружено в результате наблюдательных исследований. Мы обращаем внимание на то, что наши рекомендации не являются обязательными предписаниями для организации или

проведения исследований, выбора методологии или единого формата представления результатов.

STROBE дает общие рекомендации к оформлению отчетов об описательных наблюдательных исследованиях, а также тех исследованиях, в ходе которых изучается связь между воздействием (*exposures*) и исходами (*outcomes*) здоровья. Рекомендации STROBE относятся к трем основным типам наблюдательных исследований: когортным, «случай-контроль» и одномоментным. Авторы используют различную терминологию для описания таких исследований. Например, термины «проспективное исследование» и «продольное исследование» выступают в качестве синонимов «когортного исследования», а «исследование распространенности» — как синоним «одномоментного исследования». Мы выбрали данную терминологию, поскольку она широко используется, но часто некорректно [7] или неточно [8]. Во вставке 1 представлены отличительные признаки трех дизайнов исследования.

Сфера наблюдательных исследований

Наблюдательные исследования служат широкому спектру целей: от сообщения о полученных впервые данных в отношении потенциальной причины заболевания до проверки значимости связей (*associations*), о которых сообщалось ранее. Идеи для исследований могут возникать

Jan P. Vandembroucke¹, Erik von Elm^{2, 3}, Douglas G. Altman⁴, Peter C. Gøtzsche⁵, Cynthia D. Mulrow⁶, Stuart J. Pocock⁷, Charles Poole⁸, James J. Schlesselman^{9, 10}, Matthias Egger^{2, 11} for the STROBE Initiative

¹ Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

² Institute of Social & Preventive Medicine (ISPM), University of Bern, Bern, Switzerland

³ Department of Medical Biometry and Medical Informatics, University Medical Centre, Freiburg, Germany

⁴ Cancer Research UK/NHS Centre for Statistics in Medicine, Oxford, United Kingdom

⁵ Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark

⁶ University of Texas Health Science Center, San Antonio, United States of America

⁷ Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London, United Kingdom

⁸ Department of Epidemiology, University of North Carolina School of Public Health, Chapel Hill, United States of America

⁹ University of Pittsburgh Graduate School of Public Health, Pittsburgh, United States of America

¹⁰ University of Pittsburgh Cancer Institute, Pittsburgh, United States of America

¹¹ Department of Social Medicine, University of Bristol, Bristol, United Kingdom

Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. Translation to Russian

Much medical research is observational. The reporting of observational studies is often of insufficient quality. Poor reporting hampers the assessment of the strengths and weaknesses of a study and the generalisability of its results. Taking into account empirical evidence and theoretical considerations, a group of methodologists, researchers, and editors developed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) recommendations to improve the quality of reporting of observational studies. The STROBE Statement consists of a checklist of 22 items, which relate to the title, abstract, introduction, methods, results and discussion sections of articles. Eighteen items are common to cohort studies, case-control studies and cross-sectional studies and four are specific to each of the three study designs. The STROBE Statement provides guidance to authors about how to improve the reporting of observational studies and facilitates critical appraisal and interpretation of studies by reviewers, journal editors and readers. This explanatory and elaboration document is intended to enhance the use, understanding, and dissemination of the STROBE Statement. The meaning and rationale for each checklist item are presented. For each item, one or several published examples and, where possible, references to relevant empirical studies and methodological literature are provided. Examples of useful flow diagrams are also included. The STROBE Statement, this document, and the associated Web site (<http://www.strobe-statement.org/>) should be helpful resources to improve reporting of observational research.

Present article is Russian-language translation of the original manuscript edited by Doctor of Medicine R.T. Saygitov. Present translation was first published in Digital Diagnostics. doi: 10.17816/DD70821. It is published with minor changes related to the literary editing of the translation itself.

Keywords: STROBE, observational studies, cohort studies, case-control studies, cross-sectional studies, study design

For citation (translation): Jan P. Vandembroucke, Erik von Elm, Douglas G. Altman, Peter C. Gøtzsche, Cynthia D. Mulrow, Stuart J. Pocock, Charles Poole, James J. Schlesselman, Matthias Egger. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. Translation to Russian. *Voprosy sovremennoi pediatrii — Current Pediatrics*. 2022;21(3):173–208. (In Russ). doi: <https://doi.org/10.15690/vsp.v21i3.2426>

For citation (original article): Vandembroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, et al. (2007) Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *PLOS Medicine* 4(10): e297. <https://doi.org/10.1371/journal.pmed.0040297>

в результате клинических наблюдений или из описания биологических явлений. Неформальное видение определенных данных может также способствовать возникновению идей, ведущих к дальнейшим исследованиям. Подобно врачу-клиницисту, который осмотрел тысячи пациентов и отмечает одного, чем-то привлекшего к себе его внимание, исследователь также может заметить нечто особенное в данных. И хотя изучение одних и тех же данных с разных точек зрения может быть невозможным или нежелательным [9], для подтверждения или опровержения первоначальных наблюдений часто возникает необходимость в проведении дальнейших исследований [10]. Существующие данные могут использоваться для изучения новых предположений о потенциальных причинных факторах и вместе с тем могут быть достаточными для их опровержения или подтверждения. В некоторых случаях исследования проводятся именно с целью преодоления потенциальных проблем, связанных с предыдущими сообщениями. Главная задача таких исследований — сбор новых данных, а не анализ уже существующих. Это приводит к появлению различных точек зрения, например о преимуществах рассмотрения подгрупп или о важности заранее определенного размера выборки (*sample size*). Инициатива STROBE направлена на достижение согласованного описания наблюдательных исследований с разнообразными целями — от научных открытий до опровержения или подтверждения ранее полученных результатов. Для отдельных случаев приводятся конкретные рекомендации.

Как использовать эту статью

Эта статья связана с краткой версией рекомендаций STROBE, представленных в форме перечня и опубликованных ранее [6], и является неотъемлемой частью стандарта STROBE. Наша задача — объяснить, как правильно представлять результаты исследования, а не как его проводить. Мы прилагаем подробные пояснения к каждому пункту перечня. Пояснению предшествует пример того, что мы считаем правильно оформленным сообщением. Это не означает, что исследование, из которого был взят данный пример, было качественно представлено или выполнено. Это также не означает, что его результаты были надежными (*reliable*) и позже подтверждены другими исследователями. Это всего лишь означает, что данный конкретный фрагмент текста статьи был корректно представлен в этом исследовании. Предложенные нами пояснения и примеры мы сопроводили вставками (1–8) с дополнительной справочной информацией о некоторых теоретических моментах или технических деталях. Для более полного их понимания может потребоваться изучение учебников или методологических источников, которые нами цитируются.

Рекомендации STROBE не ориентированы на специфические исследования, такие как изучение генетических связей, моделирование инфекционных заболеваний, описание случая или серии случаев [11, 12]. Но поскольку многие ключевые пункты перечня STROBE применимы к подобным дизайнам, наши рекомендации могут оказаться уместными и полезными для тех авторов, которые представляют результаты таких исследований. Для авторов работ, посвященных изучению методов диагностики, онкомаркеров и генетических связей, могут быть полезны рекомендации STARD [13], REMARK [14] и STREGA [15].

Вставка 1. Основные дизайны исследований, охватываемые рекомендациями STROBE

Когортное исследование, исследование «случай-контроль» и одномоментное исследование представляют собой

различные подходы к изучению возникновения событий, связанных со здоровьем, в данной популяции и период времени. Эти исследования могут охватывать такие события, как возникновение болезни или наступление ремиссии, инвалидность или осложнения, смерть или выживаемость (*survival*), а также появление факторов риска.

В **когортных исследованиях** исследователи наблюдают за людьми в течение времени. Они получают данные об участниках исследования и воздействии на них изучаемых факторов на момент включения в исследование и через какое-то время оценивают результаты воздействия (исходы). Авторы обычно противопоставляют участников, которые подверглись и не подвергались воздействию, или группы людей с различными категориями воздействия. Они также могут оценивать несколько различных исходов, а кроме того, изучать переменные воздействия и исхода в разные моменты времени в ходе наблюдения. Закрытые когорты (например, возрастные) включают определенное количество участников, после чего за ними ведется наблюдение, часто на протяжении установленных промежутков времени, до фиксированной даты окончания исследования. В открытых когортах (*open cohorts*) выборка исследования (*study population*) динамична: участники исследования (например, жители города) попадают в нее и выбывают из нее в разные моменты времени. Открытая когорта может меняться в результате смертей, рождения или миграции участников исследования, но состав когорты с учетом возраста и пола может оставаться примерно одинаковым, особенно на протяжении короткого периода времени. В закрытых когортах (*closed cohorts*) могут оцениваться кумулятивная инцидентность (*cumulative incidences*) (риски) и относительная инцидентность (*incidence rates*); при сравнении групп, подвергшихся и не подвергавшихся воздействию, на основании упомянутых показателей определяется отношение рисков (*risk ratio*) или отношение частот (*rate ratio*). В открытых когортах оцениваются относительная инцидентность (*incidence rates*) и отношение относительных инцидентностей (*incidence rate ratios*).

В исследованиях «случай-контроль» авторы сравнивают эффекты воздействия между участниками с определенным исходом заболевания («случаи») и участниками без такого исхода («контроли»). Авторы стремятся отбирать «случаи» и «контроли», которые репрезентативны когорте, из которой их набирают, или происходили из поперечного «среза» популяции. Такую популяцию можно определить географически, но в более широком смысле это территория, обслуживаемая медицинскими учреждениями. Выборка «случаев» может составлять все или большую часть имеющихся «случаев», в то время как выборка «контролей» обычно составляет лишь небольшую часть лиц, не имеющих соответствующего исхода. «Контроли» представляют когорту или популяцию, в которой могут появляться «случаи». Исследователи рассчитывают отношение шансов воздействия предполагаемых причин заболевания между «случаями» и «контролями» (см. вставку 7). В зависимости от стратегии формирования групп «случаев» и «контролей», а также от особенностей изучаемой популяции отношение шансов (*odds ratio*) в исследованиях «случай-контроль» интерпретируется как отношение рисков, отношение частот (*rate ratio*) или отношение шансов для распространенности (*prevalence*) [16, 17]. Большинство опубликованных исследований «случай-контроль» представляют собой открытые когорты и, таким образом, позволяют получать прямые оценки отношения частот.

В **одномоментных исследованиях** исследователи изучают всех участников выборки в один момент времени, часто для того, чтобы выявить распространенность

воздействия, факторов риска или болезни. Некоторые одномоментные исследования являются аналитическими и нацелены на количественную оценку потенциальных причинно-следственных связей между воздействием и заболеванием. Такие исследования можно анализировать как когортное исследование, сравнивая показатели распространенности заболевания между группами воздействия, или как исследование «случай-контроль», сравнивая шансы воздействия в группах с заболеванием и без него. Сложность, которая может возникнуть в исследовании любого дизайна и которая особенно характерна для одномоментных исследований, состоит в том, чтобы установить, что воздействие предшествовало заболеванию, хотя временная последовательность воздействия и исхода иногда может быть вполне очевидной. Например, в исследовании, где переменная воздействия является врожденной или генетической характеристикой, можно быть уверенным, что воздействие предшествовало исходу, даже если мы определяем то и другое одновременно.

О пунктах перечня рекомендаций STROBE

Далее мы обсудим и объясним 22 пункта перечня рекомендаций STROBE (см. табл. 1) и приведем опубликованные примеры к каждому из них. В некоторых примерах были удалены цитаты или добавлены сокращения. 18 пунктов относятся ко всем трем типам исследований, тогда как 4 пункта касаются конкретных дизайнов. Пункты, помеченные

звездочкой (например, пункт 8*), указывают на то, что информацию следует представлять отдельно для «случаев» и «контролей» в исследованиях «случай-контроль» или для групп, подвергшихся и не подвергавшихся воздействию, в когортных и одномоментных исследованиях. Мы советуем авторам применять в своей работе все предложенные пункты, но их расположение или очередность могут быть произвольными. Например, мы предлагаем представлять результаты отдельными пунктами, признавая вместе с тем, что авторы могут объединить информацию по нескольким пунктам в одном разделе текста или в таблице.

РАЗДЕЛЫ

Название и аннотация

1А. Укажите в названии статьи или в аннотации наименование дизайна исследования, используя общепринятые термины

Пример

«Заболеваемость лейкемией среди работников обувной промышленности: исследование „случай-контроль“» [18].

Пояснение

Читатели должны иметь возможность легко идентифицировать дизайн исследования по названию статьи или аннотации. Использование общепринятых терминов способствует правильной индексации статей в электронных базах данных [19, 20].

Таблица 1. STROBE: перечень пунктов, которые должны быть представлены в отчетах о наблюдательных исследованиях

Table 1. The STROBE Statement — Checklist of Items That Should Be Addressed in Reports of Observational Studies

	№ пункта	Рекомендации
Название и аннотация	1	А. Укажите в названии статьи или в аннотации наименование дизайна исследования, используя общепринятые термины. Б. Включите в аннотацию информативное и сбалансированное обобщение того, что было сделано и обнаружено
Введение		
Актуальность/ обоснование	2	Объясните актуальность и обоснуйте необходимость проведения представляемого научного исследования
Цели	3	Укажите цели исследования, включая все предварительные гипотезы
Методы		
Дизайн исследования	4	Укажите основные элементы дизайна исследования
Условия проведения	5	Опишите условия, место и соответствующие даты проведения исследования и все связанные с этим данные, включая периоды привлечения участников, воздействия, наблюдения и сбора данных
Участники	6	А. Когортное исследование. Укажите критерии, источники и методы отбора участников. Опишите методику наблюдения. Исследование «случай-контроль». Укажите критерии отбора, источники и методы установления «случаев» и «контролей». Обоснуйте выбор «случаев» и «контролей». Одномоментное исследование. Укажите критерии, источники и методы отбора участников. Б. Когортное исследование. Для исследований с подобранными группами опишите критерии подбора и количество участников, подвергшихся и не подвергавшихся воздействию изучаемого фактора. Исследование «случай-контроль». Для исследований с подобранными группами опишите критерии подбора и количество «контролей» на один «случай»
Переменные	7	Четко определите все исходы, воздействия, прогностические факторы, потенциальные вмешивающиеся факторы, модификаторы эффекта. Опишите диагностические критерии, если применимо
Источник данных, измерения	8*	Для каждой переменной, представляющей интерес, приведите источники данных и подробные сведения о методах оценки (измерения). Укажите, были ли сопоставимы методы измерения в группах
Систематические ошибки	9	Опишите все усилия по устранению потенциальных источников систематических ошибок
Объем исследования	10	Объясните, как был определен объем (размер) исследования
Количественные переменные	11	Укажите, какие количественные переменные были использованы в анализе и, если применимо, — какие категории (группы) на основании этих переменных были выбраны и почему

Таблица 1. Продолжение
Table 1. Continuation

	№ пункта	Рекомендации
Статистические методы	12	<p>А. Опишите все статистические методы, включая использованные для контроля влияния вмешивающихся факторов.</p> <p>Б. Опишите все методы, использованные для изучения подгрупп и взаимодействий.</p> <p>В. Объясните, как устраняли последствия отсутствующих (неполных) данных.</p> <p>Г. Когортное исследование. Если применимо, опишите, как решали проблему выбывших из-под наблюдения.</p> <p>Исследование «случай-контроль». Если применимо, объясните, как проводили подбор (<i>matching</i>) «случаев» и «контролей».</p> <p>Одномоментное исследование. Если применимо, опишите аналитические методы, учитывающие стратегию формирования выборки.</p> <p>Д. Опишите использованные анализы чувствительности</p>
Результаты		
Участники	13*	<p>А. Укажите количество лиц на каждом этапе исследования: число потенциально подходящих для участия в исследовании, обследованных на предмет соответствия критериям отбора, допущенных к участию, включенных в исследование, завершивших последующее наблюдение, а также участников, чьи результаты были проанализированы.</p> <p>Б. Укажите причины выбывания на каждом этапе исследования.</p> <p>В. Рассмотрите возможность использования потоковой диаграммы</p>
Описательные данные	14*	<p>А. Охарактеризуйте участников исследования (например, укажите демографические, клинические, социальные характеристики). Приведите данные о воздействиях и потенциальных вмешивающихся факторах.</p> <p>Б. Для каждой из анализируемых переменных укажите количество участников с пропущенными данными.</p> <p>В. Когортное исследование. Опишите характеристики времени наблюдения (например, укажите среднюю и общую продолжительность наблюдения)</p>
Исходы	15*	<p>Когортное исследование. Укажите количество событий исхода или представьте сводку основных показателей исследования за время наблюдения.</p> <p>Исследование «случай-контроль». Укажите количество «случаев» по каждой категории воздействия или сводные показатели воздействия.</p> <p>Одномоментное исследование. Укажите количество событий исхода или представьте сводку основных показателей исследования</p>
Основные результаты	16	<p>А. Укажите нескорректированные оценки и, если применимо, оценки с поправкой на вмешивающиеся факторы и их точность (например, 95% доверительный интервал). Объясните, влияние каких вмешивающихся факторов было учтено и почему именно эти факторы были выбраны.</p> <p>Б. При категоризации непрерывных переменных укажите границы категорий.</p> <p>В. Если применимо, рассмотрите возможность преобразования относительного риска в абсолютный для значимого периода времени</p>
Другие анализы	17	Если проводились другие виды анализа, например анализ в подгруппах, взаимодействий, чувствительности, укажите это
Обсуждение		
Основные результаты	18	Обобщите основные результаты, ссылаясь на цели исследования
Ограничения	19	Обсудите ограничения исследования с анализом источников потенциальных систематических ошибок или неточностей. Обсудите направление действия и значимость потенциальных систематических ошибок
Интерпретация результатов	20	Дайте осторожную, но достаточно полную интерпретацию результатов с учетом целей, ограничений, многочисленных анализов, результатов схожих исследований и других уместных свидетельств
Обобщаемость	21	Обсудите обобщаемость (внешнюю достоверность) результатов исследования
Дополнительная информация		
Финансирование	22	Укажите источники финансирования и роль спонсоров в настоящем исследовании и, если применимо, в оригинальном исследовании, на котором основывается представляемая статья

Примечание. <*> — в исследованиях «случай-контроль» рекомендуется давать информацию отдельно для «случаев» и «контролей»; в когортных и одномоментных исследованиях, если применимо, — для подвергшихся и не подвергавшихся воздействию групп. В данной статье приводятся все пункты перечня рекомендаций (*checklist*) с методологическим обоснованием и примерами качественного представления исследований в опубликованных работах. Перечень рекомендаций STROBE эффективнее всего использовать в сочетании с этой статьей (имеется в свободном доступе на сайтах журналов PLoS Medicine (<http://www.plosmedicine.org/>), Annals of Internal Medicine (<http://www.annals.org/>), Epidemiology (<http://www.epidem.com/>)). Версии перечня отдельно для когортных исследований, исследований «случай-контроль» и одномоментных исследований доступны на сайте STROBE (<http://www.strobe-statement.org/>). doi: <https://doi.org/10.1371/journal.pmed.0040297.t001>

Note. <*> — give such information separately for cases and controls in case-control studies, and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies. An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Separate versions of the checklist for cohort, case-control, and cross-sectional studies are available on the STROBE Web site at <http://www.strobe-statement.org/>. doi: <https://doi.org/10.1371/journal.pmed.0040297.t001>

1Б. Включите в аннотацию информативное и сбалансированное обобщение того, что было сделано и обнаружено

Пример

«Актуальность»: ожидаемая выживаемость ВИЧ-инфицированных пациентов представляет большой интерес для общественного здравоохранения.

Цель: оценить время выживания и повозрастную смертность в когорте ВИЧ-инфицированных пациентов в сравнении с общей популяцией.

Дизайн: популяционное когортное исследование.

Условия проведения: все ВИЧ-инфицированные лица, получающие медицинскую помощь в Дании в период с 1995 по 2005 г.

Пациенты: каждому участнику Датского общенационального когортного исследования ВИЧ-инфицированных подбирали для сравнения по полу, дате рождения и месту проживания до 99 лиц из общей популяции.

Измерения. Авторы рассчитывали таблицы дожития Каплана – Мейера с возрастом в качестве временной шкалы для оценки выживаемости после 25-летнего возраста. ВИЧ-инфицированные пациенты и соответствующие им лица из общей популяции наблюдались от момента установления диагноза и до смерти, эмиграции или наступления 1 мая 2005 г.

Результаты. В исследование были включены 3990 ВИЧ-инфицированных и 379 872 лица из общей популяции. Суммарное время наблюдения составило 22 744 человеко-года (медиана — 5,8 года) и 2 689 287 человеко-лет в группе контроля (медиана — 8,4 года) соответственно. Исходы 3% участников не были отслежены. После достижения возраста 25 лет медиана выживаемости составила 19,9 года (95% ДИ 18,5–21,3) в группе ВИЧ-инфицированных и 51,1 года (ДИ 50,9–51,5) — в выборке из общей популяции. В группе ВИЧ-инфицированных выживаемость возросла до 32,5 лет (ДИ 29,4–34,7) за период с 2000 до 2005 г. При исключении лиц с гепатитом С (16%) медиана выживаемости за тот же период составила 38,9 года (ДИ 35,4–40,1). Относительный риск смерти в когорте ВИЧ-инфицированных по сравнению с таковым в выборке из общей популяции снижался с возрастом, тогда как смертность с увеличением возраста увеличивалась.

Ограничения: наблюдаемая смертность признается допустимой для анализа, если текущая максимальная длительность наблюдения превышает 10 лет.

Выводы. В эпоху высокоактивной антиретровирусной терапии оценка медианы выживаемости для молодых ВИЧ-инфицированных лиц составляет более 35 лет. Однако для дальнейшего снижения смертности ВИЧ-инфицированных в сравнении с общей популяцией по-прежнему необходимо прикладывать постоянные усилия» [21].

Пояснение

Аннотация содержит ключевую информацию, которая позволяет читателям понять исследование и принять решение о том, читать ли статью. Типичные компоненты аннотации — формулировка исследовательского вопроса, краткое описание методов и результатов, выводы [22]. В аннотации необходимо резюмировать ключевые детали исследования и включать только ту информацию, которая представлена в статье. Рекомендуется представлять ключевые результаты в цифровой форме: количество участников, оценка связей, изменчивости (*variability*) и неопределенности (*uncertainty*) (например, отношения шансов с доверительными интервалами). Мы считаем недостаточной простую констатацию наличия

или отсутствия статистически значимой связи между воздействием и исходом.

Серия заголовков, относящихся к актуальности, дизайну, проведению и анализу данных исследования, поможет читателям быстро найти интересующую их информацию [23]. Во многих журналах требуют представлять структурированные аннотации, которые, как правило, имеют более высокое качество и более информативны, чем неструктурированные резюме [24, 25].

Введение

В разделе «Введение» следует описать, почему было проведено исследование и какие вопросы и гипотезы оно затрагивает. Это позволит другим понять контекст исследования и оценить его потенциальный вклад в современные знания.

2. Актуальность/обоснование.

Объясните актуальность и обоснуйте необходимость проведения представляемого научного исследования

Пример

«Рост распространенности ожирения среди детей и подростков определил интерес к получению качественных данных о связи ожирения у детей с повышенным риском сердечно-сосудистых заболеваний и смерти во взрослом возрасте. Ожирение приводит к серьезным социальным и психологическим проблемам в детском и подростковом возрасте, однако о том, какие социальные, социально-экономические и психологические последствия оно вызывает в последующем, известно немного. Недавний систематический обзор показал, что длительные когортные исследования последствий детского ожирения, помимо последствий для физического здоровья, не проводились. Было проведено лишь два исследования социально-экономических последствий ожирения у подростков. S.L. Gortmaker и соавт. обнаружили, что жительницы США с ожирением в позднем подростковом возрасте в 1981 г. семь лет спустя реже были замужем и имели более низкий доход, чем женщины, не страдавшие ожирением в подростковом возрасте, в то время как мужчины с избыточной массой тела реже вступали в брак. J.D. Sargent и соавт. выяснили, что женщины в Великобритании (но не мужчины), страдавшие ожирением в возрасте 16 лет в 1974 г., в возрасте 23 лет зарабатывали на 7,4% меньше, чем их сверстники, не страдавшие ожирением. <...> Мы использовали данные длительного наблюдения за британской когортой 1970 года рождения для изучения социально-экономических, образовательных, социальных и психологических исходов детского ожирения во взрослом возрасте» [26].

Пояснение

Научные предпосылки исследования дают читателям необходимое представление о контексте, раскрывая суть и направление исследования, обеспечивая общее представление о том, что известно по данному вопросу и каковы пробелы в современных знаниях. В этом разделе можно сослаться на недавно проведенные исследования по схожей проблематике и систематические обзоры таких исследований.

3. Цели. Укажите цели исследования, включая все предварительные гипотезы

Пример

«Нашими основными целями было: 1) определить распространенность домашнего насилия среди пациен-

ток, обратившихся в 4 общественных центра первичной медицинской помощи, которые обслуживают пациентов с разным социально-экономическим статусом, и 2) выявить демографические и клинические различия между пациентами, которые подвергаются и не подвергаются насилию в настоящее время» [27].

Пояснение

Цели — это подробное изложение желаемого результата исследования. Тщательно продуманные цели помогают точно определить популяции, воздействия, исходы, а также другие параметры, которые будут оцениваться. Они могут быть сформулированы в форме конкретных гипотез или как вопросы, для ответа на которые и было спланировано исследование. В некоторых случаях цели могут быть менее четкими, например, на ранних фазах научного исследования. Тем не менее, в рукописи должны быть четко отражены намерения авторов. Например, если важный анализ в подгруппах или любой другой дополнительный анализ не был изначально целью исследования, но необходимость этого возникла в ходе изучения данных, это должно быть описано соответствующим образом (см. также п. 4, 17 и 20).

Методы

В разделе «Методы» рекомендуется подробно описать, что было запланировано и что было сделано, чтобы другие могли понять основные аспекты исследования, судить о том, являются ли использованные методы приемлемыми для получения надежных (*reliable*) и достоверных (*valid*) данных, а также оценить обоснованность каких-либо отклонений от первоначального плана.

4. Дизайн исследования. Укажите основные элементы дизайна исследования

Пример

«Мы использовали перекрестное сравнение случаев (*case-crossover design*), уместное в ситуациях, когда кратковременное воздействие фактора (использование телефона водителем) вызывает преходящее повышение риска редкого исхода (авария). Мы сравнивали использование мобильного телефона водителем в момент аварии и в другие подходящие для сравнения периоды. Поскольку водители выступали и в роли собственных «контролей», в исследовании учитывали их характеристики, которые могли повлиять на риск аварии и вместе с тем были неизменными в течение короткого периода времени. Ввиду того, что риски в контрольные периоды и во время аварийных поездок должны быть схожими, мы сравнивали активность использования телефона в интервале риска (время непосредственно перед аварией) и в контрольные интервалы (отрезки времени, когда участники управляли автомобилем, но не попали в аварию) в течение предыдущей недели» [28].

Пояснение

Мы советуем представлять основные элементы дизайна исследования в начале раздела «Методы» или в конце «Введения», чтобы читатели могли понять схему исследования (*study design*). Например, необходимо указать, что авторы проводили когортное исследование, где велось наблюдение за участниками в течение определенного времени, описать группу лиц, составляющих когорту, указать статус оказываемого на них воздействия. Аналогичным образом, если использован дизайн исследования «случай-контроль», должны быть описаны «случаи», «контроли» и исходная популяция (*source population*). Если проводилось одномоментное исследование, следует

описать популяцию и временную точку, в которой производили поперечный «срез». Если исследование является промежуточным вариантом трех основных дизайнов, необходимо внести дополнительную ясность. Так, в примере выше мы кратко описали принципы перекрестного сравнения случаев одного из вариантов исследования «случай-контроль» [28].

Мы рекомендуем воздержаться от простого обозначения исследования как «проспективное» или «ретроспективное», поскольку эти термины не имеют четких определений [29]. Одни используют термины «проспективный» и «когортный» как синонимы и оставляют слово «ретроспективный» для исследований «случай-контроль» [30]. Другие различают проспективные и ретроспективные когортные исследования в зависимости от периода сбора данных по отношению к тому времени, когда возникла идея исследования [31]. Третьи различают проспективные и ретроспективные исследования «случай-контроль» в зависимости от того, имелась ли информация об исследуемом воздействии в тот момент, когда отбирали подходящие «случаи» [32]. Некоторые советуют не использовать эти термины [33] или употреблять альтернативные варианты для описания когортных исследований — «параллельные» и «исторические» [34]. В рекомендациях STROBE мы не используем термины «проспективный» и «ретроспективный», впрочем, как и их альтернативы «параллельный» и «исторический». Авторам рекомендуется всякий раз при использовании данной терминологии объяснять, что конкретно имеется в виду, а самое главное — точно описывать, как и когда происходил сбор данных.

В первой части раздела «Методы» можно также указать, является ли рукопись частью серии публикаций по результатам исследования. Если новая рукопись соответствует первоначальным целям этого исследования, на это обычно указывают ссылкой на более раннюю публикацию и краткое повторение ее основных особенностей. Однако цели исследования могут меняться со временем.

Авторы часто используют данные в целях, для которых они изначально не предназначались, включая в исследование, например, официальную статистику естественного движения населения, которую первоначально собирали в административных целях, анкетные данные, которые изначально были включены лишь для полноты информации, или образцы крови, которые собирали в связи с другими причинами. Например, исследование Physicians' Health Study (рандомизированное контролируемое исследование аспирина и каротина) впоследствии было использовано для демонстрации того, что точечная мутация гена фактора V ассоциирована с высоким риском венозного тромбоза, но не инфаркта миокарда или инсульта [35]. Вторичное использование имеющихся данных — творческая часть наблюдательного исследования, которое не обязательно делает полученные результаты менее убедительными или менее важными. Тем не менее, краткое повторение первоначальных целей поможет читателям понять контекст исследования и возможные ограничения в использовании данных.

5. Условия проведения. Опишите условия, место и соответствующие даты проведения исследования, включая периоды привлечения участников, воздействия, наблюдения и сбора данных

Пример

«В исследование Pasitos Cohort Study включали беременных женщин, наблюдавшихся в клиниках „Мать, младенец и дитя“ в городах Сокорро и Сан-Элизиарио округа Эль-Пасо (штат Техас, США), а также в клинике „Мать

и дитя” Мексиканского института социального обеспечения в городе Сьюдад-Хуарес (Мексика) с апреля 1998 по октябрь 2000 г. Изначально, до рождения детей, женщин, включенных в когорту, расспрашивали о домашней обстановке. В этом продолжающемся когортном исследовании наблюдения проводили с полугодовой периодичностью, начиная с достижения новорожденными возраста 6 мес» [36].

Пояснение

Читателям необходимо знать условия и место проведения исследования, чтобы оценить контекст и обобщаемость его результатов. Такие воздействия, как средовые факторы и терапевтические вмешательства, могут со временем меняться, как и методы исследования. Информация о времени проведения исследования, наборе участников и наблюдении за ними помещает исследование в исторический контекст, что важно для интерпретации результатов.

Информация об условиях проведения включает место набора пациентов или источник информации о них (например, список избирателей, поликлиника, система регистрации онкологических заболеваний или центр третичной медицинской помощи). К информации о месте проведения исследования относятся сведения о стране, городе, госпитале или медицинской практике, где проводилось исследование. Мы советуем указывать даты, а не только длительность периодов времени. Могут быть разные даты для воздействия, возникновения заболевания, набора участников, начала и конца периода наблюдения, сбора данных. Следует отметить, что почти 80% из 132 статей, опубликованных в онкологических медицинских журналах, в которых были представлены результаты анализа выживаемости, включали даты начала и окончания набора пациентов, и лишь в 24% статей сообщали дату окончания наблюдения [37].

6. Участники

6А. Когортное исследование. Укажите критерии, источники и методы отбора участников.

Опишите методику наблюдения

Пример

«Участницы исследования Iowa Women’s Health Study были случайной выборкой из числа всех женщин в возрасте 55–69 лет, получивших автомобильные права в штате Айова в 1985 г., что составило примерно 94% всех женщин штата Айова этой возрастной группы. <...> Вопросники для определения жизненного статуса и изменения адреса проживания рассылали по почте в октябре 1987 и августе 1989 г. <...> Заболеваемость раком (за исключением немеланомного рака кожи) устанавливали по данным Регистра здравоохранения штата Айова. <...> Для участниц Iowa Women’s Health Study из регистра подбирали пары, сопоставимые по имени, настоящей фамилии и девичьей фамилии, почтовому коду и дате рождения» [38].

6А. Исследование «случай-контроль».

Укажите критерии отбора, источники и методы установления «случаев» и «контролей».

Обоснуйте выбор «случаев» и «контролей»

Пример

«Случаи меланомы кожи, диагностированные в 1999–2000 гг., выявляли по данным Регистра рака штата Айова. <...> Контролем по данным этого же регистра были случаи колоректального рака, диагностированные в тот же период. Пациентов с колоректальным раком использовали в качестве «контролей», поскольку это рас-

пространенное заболевание с высокой выживаемостью, а воздействие мышьяка не связано с развитием этого заболевания» [39].

6А. Одномоментное исследование. Укажите критерии, источники и методы отбора участников

Пример

«Ретроспективно выявляли пациентов с основным диагнозом „инфаркт миокарда“ (код 410) в соответствии с МКБ-9 по кодам диагнозов при выписке, за исключением кодов с пятой цифрой 2, которые обозначают последующий эпизод лечения. <...> Из числа всех пациентов с диагнозом „инфаркт миокарда“, получавших медицинскую помощь по программе Medicare с февраля 1994 по июль 1995 г., была сформирована случайная выборка. <...> Критериями включения были госпитализация более чем через 30 мин, но менее чем через 12 ч после появления болей в груди, с подъемом сегмента ST не менее 1 мм в двух последовательных отведениях на электрокардиограмме при поступлении» [40].

Пояснение

Детальное описание участников исследования помогает читателям понять применимость (*applicability*) результатов. Исследователи обычно ограничивают целевую популяцию (*study population*) с помощью ряда клинических, демографических и других характеристик. Типичные критерии отбора (*eligibility criteria*) — возраст, пол, диагноз, сопутствующие заболевания. Несмотря на их важность, критерии отбора часто не описывают должным образом. Так, критерии отбора участников не были указаны в 17 из 49 статей (35%) с результатами наблюдательных исследований инсульта [5].

Критерии отбора в исследовании могут быть описаны как критерии включения (*inclusion criteria*) и исключения (*exclusion criteria*), хотя такое разделение не всегда необходимо и полезно. Мы советуем авторам описать все критерии отбора, а также указать, из какой группы отбирали участников для исследования (например, общая популяция жителей региона или страны), и описать метод отбора (например, направление участников клиницистами или их самостоятельная регистрация по объявлению о наборе в исследование).

Знание деталей процедуры наблюдения (действия по минимизации доли не ответивших или выбывших из-под наблюдения участников, одинаковость процедуры наблюдения для всех участников) позволяет сделать выводы о достоверности (*validity*) результатов. Например, если в исследовании для выявления острой инфекции использовали антитела IgM, необходимо указать интервал между их определениями в крови, что позволит читателям оценить, не были ли некоторые случаи инфекции пропущены из-за слишком длительного интервала между анализами [41]. В исследованиях, где процедуры наблюдения различались в группах, подвергшихся и не подвергавшихся воздействию, читатели могут распознать существенную систематическую ошибку, возникшую из-за различий в установлении событий или разницы в числе не ответивших или выбывших из-под наблюдения участников [42]. Поэтому мы советуем описывать методы наблюдения за участниками и указывать, были ли они одинаковы для всех участников и позволяют ли описать изучаемые переменные достаточно полно (см. п. 14).

В исследованиях «случай-контроль» выбор «случаев» и «контролей» имеет решающее значение для интерпретации результатов, а метод их отбора представляет первостепенную важность для определения достоверности

исследования (*study validity*). В целом «контроли» должны отражать популяцию, из которой происходят «случаи». Для подбора «контролей» используют различные методы, каждый из которых имеет свои преимущества и недостатки. Для «случаев», происходящих из общей популяции, это выборка из реестра популяции, случайный обзвон, привлечение соседей или друзей. «Контроли», подобранные из соседей или друзей, могут помочь лучше оценить воздействие, поскольку позволяют скорректировать влияние других факторов риска, общих для «случаев» и «контролей» [17]. Подбор «контролей» из лиц, страдающих другими заболеваниями, может иметь преимущества перед отбором из популяции, в частности, для «случаев», связанных с госпитализацией, поскольку он проще в исполнении, лучше отражает госпитальную популяцию, отличается сопоставимым откликом «случаев» и «контролей» при включении в исследование. Однако здесь могут возникать проблемы, если изучаемое воздействие повышает риск развития заболевания или связанной с ним госпитализации в группе «контролей» [43, 44]. Для решения этой проблемы часто используют несколько наиболее подходящих для сравнения болезней [45].

6Б. Когортное исследование. Для исследований с подобранными группами опишите критерии подбора и количество участников, подвергшихся и не подвергавшихся воздействию изучаемого фактора

Пример

«Для каждого пациента, изначально получавшего статин, в качестве контроля подбирали пару — участника, не получавшего статин, в соответствии со следующим протоколом. Сначала для каждого пациента в общей когорте на основании обширного списка факторов рассчитывали показатель вероятности использования статинов или риска возникновения сепсиса. На втором этапе каждому пациенту, получающему статин, подбирали небольшой пул не получающих статин и сопоставимых по полу, возрасту (плюс-минус 1 год) и указателю даты (плюс-минус 3 мес). На третьем этапе из этого пула выбирали „контроль“, наиболее близкий по показателю вероятности (в пределах 0,2 величины стандартного отклонения) для каждого „случая“ в соотношении 1 : 1. Контроли, оставшиеся без пары, исключали из исследования» [46].

6Б. Исследования «случай-контроль». Для исследований с подобранными группами опишите критерии подбора и количество «контролей» на один «случай»

Пример

«Мы планировали отобрать по 5 „контролей“ для каждого „случая“ из числа индивидуумов исследуемой популяции, не имеющих записи о диагнозе „аутизм“ или других расстройствах аутистического спектра (РАС) в медицинской карте врача общей практики, живых на момент проведения исследования и зарегистрированных в том же учреждении, что и „случаи“ с диагнозом РАС. „Контроли“ подбирали к „случаям“ индивидуально, по году рождения (не более чем на год младше или старше), полу и общей практике. Для каждого из 300 „случаев“ подбирали 5 „контролей“, отвечающих всем критериям подбора. Оставшиеся 994 „контроля“ были исключены из исследования...» [47].

Пояснение

Подбор пар (*matching*) наиболее характерен для исследований «случай-контроль», но порой его используют и в когортных исследованиях для формирования групп,

сопоставимых в начале наблюдения. Использование метода подбора пар в когортных исследованиях позволяет получить группы, напрямую сопоставимые по потенциальным вмешивающимся факторам (*confounders*), и вызывает меньше сложностей, чем в исследованиях «случай-контроль». Например, нет необходимости учитывать переменные, использованные для подбора пар, при оценке относительного риска [48]. Поскольку применение метода подбора пар в когортном исследовании повышает статистическую точность, исследователи могут использовать его при анализе данных и таким образом получить более узкие доверительные интервалы.

В исследованиях «случай-контроль» подбор пар проводят для повышения эффективности исследования (*study's efficiency*), поскольку это помогает добиться схожего распределения переменных в группах «случаи» и «контроли», в частности распределения потенциальных вмешивающихся факторов [48, 49]. Поскольку подбор пар может быть выполнен различными способами, с одним или более «контролями» для каждого «случая», следует обосновать выбор переменных для сопоставления и описать детали используемого метода. Часто используемыми методами подбора пар являются частотное сопоставление (также называемое групповым сопоставлением; *frequency/group matching*) и индивидуальное сопоставление (*individual matching*). В случае частотного сопоставления «контроли» выбирают так, чтобы распределение переменных, используемых для подбора пар, стало идентичным или подобным их распределению в группе «случаев». Индивидуальное сопоставление означает поиск одного или нескольких «контролей» для каждого «случая». Хотя метод подбора пар интуитивно кажется привлекательным, а иногда — полезным, сопоставление в исследованиях «случай-контроль» имеет множество недостатков и не всегда уместно, что необходимо учитывать при анализе (см. вставку 2).

Вставка 2. Подбор пар в исследованиях «случай-контроль»

При проведении исследований «случай-контроль» необходимо продумать, сопоставлять ли «контроли» со «случаями», и если да, то по каким переменным, насколько точный метод сопоставления использовать и какой метод статистического анализа применять. Отказ от подбора пар может привести к тому, что распределение некоторых ключевых вмешивающихся факторов (например, возраст, пол) будет радикально отличаться в группах «случаи» и «контроли». Хотя эти различия можно скорректировать в ходе анализа, статистическая эффективность исследования может значительно снизиться.

Использование процедуры подбора пар в исследованиях «случай-контроль» и интерпретация ее результатов сопряжены с трудностями, особенно если подбор пар проводится по нескольким факторам риска, некоторые из которых могут быть связаны с воздействием, представляющим наибольший интерес [50, 51]. Например, при изучении влияния оральных контрацептивов на возникновение инфаркта миокарда исходя из фармако-эпидемиологической базы данных с информацией о тысячах женщин, которые могут быть потенциальными «контролями», у исследователей может возникнуть соблазн выбрать «контроли» со значениями факторов риска, схожими с таковыми в каждом случае инфаркта миокарда. Одна из задач — учесть факторы, которые могут повлиять на назначение оральных контрацептивов и таким образом устранить *искажение показаний* (*confounding by indication*; здесь: искажение оценки влияния лечения на исход болезни при наличии показа-

ния или противопоказания к данному лечению, которое также связано с исходом). Однако в результате будет сформирована *нерепрезентативная* (для исходной популяции женщин, получающих оральные контрацептивы) контрольная группа, поскольку «контроли» будут старше лиц в исходной популяции (*source population*) по причине того, что пациенты с инфарктом миокарда, как правило, старшего возраста. Это может иметь ряд последствий. Простой анализ данных выявит отношения шансов, которые обычно смещены в сторону единицы, если фактор, использованный для подбора пар, связан с воздействием. Выход из данной ситуации — сопоставление «случаев» и «контролей» или проведение стратифицированного анализа (см. п. 12Г). Кроме того, поскольку подобранная контрольная группа перестает быть репрезентативной для популяции в целом, распределение фактора воздействия среди «контролей» больше не может использоваться для оценки дополнительного популяционного риска (*population attributable fraction*) (см. вставку 7) [52]. К тому же влияние фактора, используемого для подбора пар, уже не может быть изучено, а поиск хорошо подобранных «контролей» может быть обременительным. В таких случаях дизайн с использованием контрольной группы как есть (сформированной без применения процедур подбора пар. — Прим. ред.) более предпочтителен, поскольку получить такие «контроли» легче, а сама контрольная группа будет больше. *Избыточный подбор* (*overmatching*) — еще одна проблема, способная снизить эффективность исследований «случай-контроль», в которых применяется процедура подбора пар, а в некоторых ситуациях — стать причиной систематических ошибок. Информация теряется, а эффективность исследования снижается, если переменная сопоставления тесно связана с фактором воздействия. В результате многие участники в сопоставляемых группах будут, как правило, иметь одинаковые или схожие уровни воздействия, и, следовательно, при сравнении не будет получена необходимая информация. Подбор пар способен внести некорректируемую систематическую ошибку в том случае, если переменная, применяемая для подбора пар, не будет вмешивающимся фактором, но будет определять причинно-следственную связь между воздействием и заболеванием. Например, экстракорпоральное оплодотворение связано с повышенным риском перинатальной смерти вследствие увеличения числа многоплодных родов и рождения младенцев с низкой массой тела [53]. Сопоставление по многоплодию или массе тела при рождении приведет к смещению результатов в сторону нуля, и такие данные нельзя будет скорректировать в ходе анализа.

Хотя интуитивно подбор пар кажется привлекательным методом формирования сравниваемых групп, методологи не рекомендуют прибегать к нему в исследованиях «случай-контроль» без должного обоснования. Они советуют тщательно и продуманно рассмотреть каждый потенциальный фактор сопоставления с пониманием того, что он может быть измерен и использован для статистической коррекции, а не с целью подбора пар. В результате можно добиться сокращения числа используемых факторов сопоставления, более широкого использования частотного сопоставления, что позволяет избежать некоторых проблем, рассмотренных выше, а также проведения большего числа исследований «случай-контроль» без какого-либо сопоставления вообще [54]. Подбор пар остается наиболее предпочтительным или даже необходимым методом, если распределение вмешивающегося фактора (например, возраста) в сравниваемых группах существенно отличается [48, 49].

Даже очень простая процедура подбора пар может быть плохо описана. Например, авторы могут утверждать, что «контроли» подбирали к «случаям» «в пределах 5 лет» или «используя 5-летний диапазон». Означает ли это, что для «случая» в возрасте 54 лет возраст «контроля» должен быть от 50 до 54 лет? Или от 49 до 59? А если используется более широкий временной промежуток (например, 10 лет), возникает опасность остаточного влияния (*residual confounding*) возраста как вмешивающегося фактора (см. вставку 4), например потому что «контроли» будут в среднем моложе, чем «случаи».

7. Переменные. Четко определите все исходы, воздействия, прогностические факторы, потенциальные вмешивающиеся факторы, модификаторы эффекта. Опишите диагностические критерии, если применимо

Пример

«В анализ включали только случаи серьезных врожденных пороков развития. Незначительные аномалии, определяемые согласно списку исключений Европейского регистра врожденных аномалий (EUROCAT), исключали из рассмотрения. Если у ребенка имелось более 1 врожденного порока развития в одной системе органов, эти пороки считали как 1 исход при анализе по системам органов. <...> При статистическом анализе возраст беременной в момент родов и число предыдущих родов рассматривали в качестве потенциально вмешивающихся факторов. Возраст матери в момент отмены противозачаточных средств и в момент родов рассценивали как потенциальные модификаторы эффекта» [55].

Пояснение

Авторы должны дать определение всем переменным, рассматриваемым и включенным в анализ, в том числе исходам, воздействиям, факторам риска, потенциальным вмешивающимся факторам и модификаторам эффекта (факторы, изменяющие эффект изучаемого воздействия). Исходы заболевания необходимо достаточно подробно описать в соответствии с диагностическими критериями. Это относится к критериям отбора «случаев» в исследовании «случай-контроль», к случаям заболевания, диагностированным во время наблюдения в когортных исследованиях, к заболеваниям, изучаемым в одномоментных исследованиях. Четкие определения и шаги, предпринятые для их соблюдения, особенно важны в случае заболевания, представляющего первичный интерес для исследования.

Для некоторых исследований уместными терминами для описания воздействия могут быть «детерминанта» или «предиктор», а для описания исходов — «конечные точки» (*endpoints*). В многомерных моделях авторы иногда используют понятие «зависимые переменные» (*dependent variable*) для описания исходов и «независимые» (*independent variable*) или «объясняющие переменные» (*explanatory variable*) — для воздействия и вмешивающихся факторов. Последнее не является точным, поскольку не позволяет отличить воздействие от вмешивающегося фактора.

Если измеряется большое количество переменных, которые затем включаются в предварительный анализ на ранней фазе исследования, обдумайте возможность создания списка с детальным описанием каждой переменной в приложении, в дополнительной таблице или отдельной публикации. Следует отметить, что в журнале *International Journal of Epidemiology* недавно ввели новый раздел «Профили когорты», который включает подробную информацию о том, какие измерения

проводились в различные моменты времени в конкретных исследованиях [56, 57]. Мы советуем авторам перечислять все переменные, рассматриваемые в ходе статистического анализа, а не описывать только те переменные, которые вошли в окончательную модель (см. п. 16А) [58, 59].

8. Источники данных, измерения. Для каждой переменной, представляющей интерес, приведите источники данных и подробные сведения о методах оценки (измерения). Укажите, были ли сопоставимы методы измерения в группах

Пример 1

«Общее потребление кофеина рассчитали, используя в основном данные Министерства земледелия США о составе продуктов. В соответствии с этими расчетами предполагалось, что в чашке кофе содержится 137 мг кофеина, в чашке чая — 47 мг, в банке или бутылке газированного напитка на основе ореха колы — 46 мг, в порции шоколадных конфет — 7 мг. Этот метод измерения потребления кофеина был признан достоверным как при изучении когорты NHS I, так и в похожем исследовании когорты мужчин — медицинских работников. <...> Выявление артериальной гипертензии на основании сообщаемой пациентом информации было признано надежным методом установления заболевания согласно данным когорты NHS I» [60].

Пример 2

«Образцы для анализов, полученные от подобранных в пары „случаев“ и „контролей“, всегда исследовали вместе одной партией, и персонал лаборатории не имел возможности отличить образцы пациентов группы „случаев“ от группы „контролей“» [61].

Пояснение

Способ измерения воздействия, вмешивающихся факторов и исходов влияет на надежность (*reliability*) и достоверность (*validity*) исследования. Ошибки измерения и классификации воздействия или исходов могут затруднить выявление причинно-следственных связей или привести к появлению ложных связей. Ошибки измерения потенциальных вмешивающихся факторов могут увеличить риск остаточного (сохраняющегося после коррекции) искажения (*residual confounding*) [62, 63]. Поэтому если авторы сообщают о результатах любых исследований достоверности или надежности оценок или измерений, полезно включить в описание характеристики референсных стандартов, если таковые использовались. Вместо того чтобы просто процитировать исследование, установившее достоверность методики (как это сделано в первом примере), мы советуем авторам представить оценку достоверности и надежности, что может в дальнейшем использоваться для корректировки ошибок измерения или при анализе устойчивости результатов исследования к изменениям исходных условий (*sensitivity analyses*) (см. п. 12Д и 17).

Кроме того, важно знать, различались ли сравниваемые группы по способу сбора данных. Это может быть существенно при лабораторной оценке (как во втором примере) и в других ситуациях. Например, если при интервьюировании опрашивать сначала все «случаи», а потом все «контроли» (или наоборот), возможно возникновение систематической ошибки в результате обучения интервьюеров (*learning curve*). Решение этой проблемы — рандомизация порядка опроса. Риск систематиче-

ской ошибки информации возрастает, если в сравниваемых группах проводили неодинаковые диагностические тесты или в одной группе проводили больше тестов того же типа, что и в другой (см. п. 9).

9. Систематические ошибки. Опишите все усилия по устранению потенциальных источников систематических ошибок

Пример 1

«В большинстве исследований „случай-контроль“, посвященных изучению суицида, в контрольные группы включали живых участников, мы же решили сформировать контрольную группу из умерших вследствие других причин. <...> В качестве источников информации о факторах риска в контрольной группе выступали лица, недавно потерявшие члена семьи или близкого человека. Эти источники более сопоставимы с источниками информации, используемыми при изучении суицида, чем контрольная группа из живых лиц» [64].

Пример 2

«Систематическая ошибка выявления может влиять на связь между сахарным диабетом 2-го типа (СД2) и первичной открытоугольной глаукомой (ПОГ), если женщины с СД2 находятся под более пристальным контролем офтальмолога, чем женщины без такого диагноза. Мы сравнили среднее число обследований зрения у женщин, страдающих и не страдающих СД2. Также пересчитали относительный риск развития ПОГ с учетом ковариат, ассоциированных с более тщательным наблюдением за состоянием органов зрения (сообщения пациентов о признаках катаракты, макулярной дегенерации, частота проверок зрения и медицинских осмотров)» [65].

Пояснение

Систематические ошибки (*bias*) в исследованиях приводят к результатам, отличающимся от истины (см. вставку 3). Читателям важно знать, какие меры принимались во время проведения исследования для уменьшения возможных систематических ошибок. В идеале исследователи должны тщательно оценить возможные источники систематических ошибок еще на этапе планирования исследования. На стадии написания отчета мы рекомендуем авторам всегда оценивать вероятность наличия соответствующих систематических ошибок. В частности, следует обсудить и по возможности оценить направление и величину смещения. Например, в исследованиях «случай-контроль» могут возникать систематические ошибки информации (измерения; *information bias*), но их риск можно уменьшить, тщательно отбирая подходящую контрольную группу (см. пример 1) [64]. Различия в медицинском наблюдении за участниками — проблема, описанная в примере 2 [65]. Следовательно, авторы приводят более подробные сведения, описывают дополнительные данные, собранные для того, чтобы решить эту проблему. Если авторы использовали программы контроля качества на этапе сбора данных для предотвращения возможных отклонений (*“drift”*) при измерении переменных в длительных когортных исследованиях (*longitudinal studies*) или для того, чтобы свести к минимуму вариабельность при участии нескольких наблюдателей, это следует описать.

К сожалению, авторы зачастую не учитывают серьезные систематические ошибки, когда представляют результаты исследований. Среди 43 исследований («случай-контроль» и когортные), опубликованных в период 1990–1994 гг., в которых изучался риск повторного

возникновения рака у пациентов с онкологическими заболеваниями в анамнезе, систематические ошибки медицинского наблюдения (*medical surveillance bias*) упоминались лишь в 5 статьях [66]. Анализ отчетов об исследованиях психического здоровья, опубликованных в течение 1998 г. в трех журналах по психиатрии, показал, что только в 13% статей из 392 сообщали о систематических ошибках отклика (различия в характеристиках тех, кто вызвался принять участие в исследовании, и тех, кто отказался; *response bias*) [67]. Обзор когортных исследований, посвященных изучению инсульта, показал, что авторы 14 из 49 статей (28%), опубликованных в период 1999–2003 гг., упоминали о потенциальной систематической ошибке отбора (*selection bias*) участников исследования, а в 35 публикациях (71%) допускалась возможность того, что любой тип систематической ошибки мог повлиять на результаты [5].

Вставка 3. Систематическая ошибка

Систематическая ошибка (смещение; *bias*) — это однонаправленное отклонение результата исследования от истинного значения. Как правило, такие смещения возникают в результате ошибок планирования или проведения исследования и в последующем не могут быть исправлены. Систематическая ошибка и конфаундинг (влияние вмешивающихся факторов; *confounding*) не являются синонимами. Систематические ошибки возникают в результате использования неверной информации или ошибок в отборе участников исследования, что приводит к ложным ассоциациям. Конфаундинг порождает связи, которые действительно существуют, но не могут быть интерпретированы как причинно-следственные, поскольку в основе этих связей — некоторые неучтенные факторы, ассоциированные и с фактором воздействия, и с анализируемым исходом (см. вставку 5). Кроме того, следует отличать систематическую ошибку от случайной ошибки (*random error*) — отклонения от истинного значения, вызванного статистическими колебаниями (в любом направлении) в результатах измерений. Описаны многие возможные источники систематических ошибок, при этом используются разные термины [68, 69]. Мы предлагаем к рассмотрению две простые категории ошибок: ошибки информации (*information bias*) и ошибки отбора (*selection bias*).

Ошибка информации возникает в связи с систематическими различиями в полноте или точности данных, что приводит к неодинаковой в сравниваемых группах классификации (*differential misclassification*) индивидуальных в отношении факторов воздействия или исходов. Например, если женщины, страдающие диабетом, проходят более регулярные и тщательные офтальмологические обследования, выявление глаукомы будет более полным, чем у женщин, не страдающих диабетом (см. п. 9) [65]. Пациенты, принимающие лекарство, вызывающее неспецифическую диспепсию, могут проходить гастроскопию чаще, и, как результат, в этой группе обнаруживается большее количество язв желудка, чем у пациентов, не принимающих такое лекарство, даже если этот препарат не способствует язвообразованию. Этот тип ошибки информации также называется «ошибкой выявления» (*detection bias*) или «ошибкой медицинского наблюдения» (*medical surveillance bias*). Один из способов оценить ее влияние — измерить интенсивность медицинского наблюдения в различных исследуемых группах и внести поправки в статистический анализ. В исследованиях «случай-контроль» ошибка информации возникает, если «случаи» более или менее

точно вспоминают прошлые воздействия по сравнению с «контролями», не страдающими таким заболеванием, или если они более или менее охотно сообщают о них (т.н. «ошибка памяти»; *recall bias*). «Ошибка интервьюера» (*interviewer bias*) может возникнуть, если интервьюеры осведомлены о гипотезе исследования и подсознательно или сознательно собирают данные выборочно [70]. В таких случаях полезно прибегать к процедуре «ослепления» (*blinding*) как участников исследования, так и исследователей.

Ошибка отбора может возникать в исследованиях «случай-контроль», если вероятность отбора «случаев» или «контролей» связана с изучаемым воздействием. Например, врач, набирающий участников для исследования тромбоза глубоких вен, может диагностировать это заболевание у пациентки, которая жалуется на боли в ногах и принимает оральные контрацептивы, а у женщины с аналогичными жалобами, не принимающей такие препараты, тромбоз глубоких вен может быть не диагностирован. Подобных ошибок можно избежать, подбирая «случаи» и «контроли» из числа тех, кто прошел одинаковое диагностическое обследование [71].

Аналогичным образом ошибки отбора могут возникать при использовании данных регистров пациентов. В частности, если известна возможная связь между воздействием и развитием заболевания, «случаи» могут с большей вероятностью быть внесены в регистр, если они подверглись воздействию предполагаемого этиологического фактора [72]. «Ошибка отклика» (*response bias*) — еще один тип ошибки отбора, который возникает, если различия в характеристиках между теми, кто отвечает, и теми, кто отказывается от участия в исследовании, влияют на оценку распространенности (*prevalence*), инцидентности (*incidence*) и иногда — связей. В целом ошибка отбора влияет на внутреннюю валидность (*internal validity*) исследования, тогда как обычные проблемы, которые могут возникнуть при наборе участников в исследование, влияют на внешнюю (*external validity*), а не на внутреннюю валидность исследования (см. п. 21).

10. Объем исследования. Объясните, как был определен объем (размер) исследования.

Пример 1

«Количество случаев в регионе в период исследования определило объем выборки» [73].

Пример 2

«Исследование случаев послеродовой депрессии в регионе выявило, что распространенность этого состояния составляет 19,8%. Предположив, что депрессия возникает у 20% матерей, родивших детей с нормальной массой тела, при отношении шансов, равном 3 для матерей маловесных детей, пришли к выводу о необходимости набора 72 участников в исследование „случай-контроль“ (1 „контроль“ на 1 „случай“) для достижения мощности исследования 80% и статистической значимости 5%» [74].

Пояснение

Исследование должно быть достаточно крупным, чтобы получить точечную оценку с достаточно узким доверительным интервалом (*confidence interval*) и убедительно ответить на вопрос, поставленный авторами работы. Большой размер выборки необходим, чтобы отличить слабые связи от отсутствия таковых. Небольшие исследования часто дают ценную информацию, но широкие доверительные интервалы могут указывать на то, что эта

информация вносит меньший вклад в современные знания по сравнению с исследованиями, предоставляющими оценки с более узкими доверительными интервалами. Кроме того, небольшие исследования, демонстрирующие «интересные» или «статистически значимые» связи, публикуются чаще, чем такого же размера исследования без каких-либо «значимых» результатов. Хотя такие работы могут сигнализировать о первых шагах к важному открытию, читатели должны быть осведомлены об их потенциальных недостатках.

Важность определения необходимого размера выборки (*sample size*) в наблюдательных исследованиях зависит от контекста. Если осуществляется анализ данных, которые уже доступны и были получены для других целей, основной вопрос — даст ли анализ данных результаты с достаточной статистической точностью (*statistical precision*), чтобы внести существенный вклад в литературу. Соображения относительно размера выборки в таких случаях не имеют принципиального значения. Формально *предварительный* расчет размеров выборки может быть полезен, когда планируется новое исследование [75, 76]. Такие расчеты связаны с большей неопределенностью, чем предполагает одно число, которое обычно указывается. Например, оценки частоты наступления интересующего события или другие предположения, важные для расчетов, обычно неточны, если не сказать гипотетичны [77]. Точность, достигаемая при окончательном анализе, не может быть спрогнозирована заранее, так как может снижаться в результате включения вмешивающихся факторов в многомерный анализ [78] вместе с точностью измерения ключевых переменных [79] и при исключении из анализа отдельных участников.

Лишь немногие эпидемиологические исследования содержат данные о размерах выборки [4, 5]. Мы поддерживаем авторов в их стремлении описать надлежащим образом расчет размера выборки, если таковой проводили. В других ситуациях уместно будет указать соображения, определившие размер выборки (например, доступная выборка фиксированных размеров, как в примере 1, упомянутом выше). Если наблюдательное исследование было прервано до получения статистически значимых результатов, читателям необходимо сообщить об этом. Не стоит запугивать читателей обоснованием размера исследования (*study size*), выполненным после его завершения, или ретроспективными расчетами его мощности [77]. С точки зрения читателя, доверительный интервал хорошо описывает достигнутую в итоге статистическую точность исследования. Необходимо понимать, что доверительные интервалы отражают лишь статистическую неопределенность, а не всю неопределенность, которая может присутствовать в исследовании (см. п. 20).

11. Количественные переменные.

Укажите, какие количественные переменные были использованы в анализе и, если применимо, — какие категории (группы) на основании этих переменных были выбраны и почему

Пример

«Пациенты с оценкой по шкале комы Глазго (ШКГ) менее 8 баллов считаются тяжелотравмированными. Оценка 9 баллов и более говорит о менее опасных повреждениях мозга. Мы оценивали связь оценки по ШКГ этих двух категорий с наступлением смерти пациентов в течение 12 мес после повреждения» [80].

Пояснение

Именно исследователи выбирают, как собирать и анализировать количественные данные о воздействии, модификаторах эффекта и вмешивающихся факторах. Например, они могут разбить область значений непрерывной переменной воздействия (*continuous exposure variable*) на несколько интервалов, создав тем самым новую, категориальную переменную (*categorical variable*) (см. вставку 4). Выбор способа группировки данных может оказать существенное влияние на последующий анализ [81, 82]. Советуем авторам объяснять, почему и как группировали количественные данные, в том числе указывать количество категорий, точки разделения (*cut-points*), среднее арифметическое значение или медиану для каждой из групп. Если данные описываются в форме таблицы, для каждой категории необходимо указать число «случаев», «контролей», лиц, подвергшихся воздействию (*persons at risk*), человеко-время для периода воздействия (*person-time at risk*). Таблицы не должны включать лишь оценки эффективности или результаты подбора модели.

Исследователи могут анализировать воздействие как непрерывное, чтобы сохранить максимум информации. При этом необходимо понимать характер влияния воздействия на исход. Поскольку это влияние может быть по ошибке автоматически принято как линейное, все отклонения от линейной зависимости должны быть изучены. Необходимо упомянуть альтернативные модели, применимость которых для анализа авторы изучали (логарифмическое преобразование, возведение в квадрат или сглаживание). Существуют несколько методов для анализа связей в случае нелинейной зависимости между воздействием и исходом [82–84]. Кроме того, повысить информативность исследования поможет одновременное представление результатов анализа и непрерывных и сгруппированных данных количественного показателя воздействия, представляющего основной интерес.

В недавно опубликованном обзоре было показано, что в двух третях статей по эпидемиологии изучали количественные переменные воздействия (*quantitative exposure variables*) [4]. В 42 статьях из 50 (84%) переменные были сгруппированы в несколько упорядоченных категорий, но часто без какого-либо обоснования сделанного выбора. В 15 статьях сообщалось об использовании линейных ассоциаций для моделирования непрерывного воздействия, но только 2 из них содержали результаты проверки на линейность. В другом обзоре литературных источников по психологии дихотомия была обоснована в 22 статьях из 110 (20%) [85].

Вставка 4. Группировка данных

Непрерывные данные могут быть разбиты на интервалы (*grouping*) по нескольким причинам [86]. При сборе данных, возможно, лучше использовать порядковую переменную (*ordinal variable*), чем искать искусственно созданную непрерывную меру воздействия, основанную на воспоминаниях за несколько лет. Категории также могут быть уместны в презентациях, например, для представления всех переменных в одном стиле или для демонстрации зависимости «доза-ответ».

Данные можно группировать для упрощения анализа, к примеру чтобы избежать предположения о линейности. Однако при группировании данных теряется информация, следовательно, может снизиться статистическая мощность (*statistical power*) [87], особенно если авторы прибегают к дихотомии [82, 85, 88]. Группировка значений

вмешивающегося фактора может привести к появлению остаточного искажения (*residual confounding*), в результате которого влияние этой переменной останется нескорректированным (см. вставку 5) [62, 89]. Увеличение числа категорий может уменьшить потери статистической мощности и остаточное искажение, что особенно уместно в крупных исследованиях. В небольших исследованиях могут использоваться несколько групп из-за ограниченного количества данных.

Исследователи могут определять точки разделения (*cut-points*) для выделения категорий на основе значений, часто используемых в диагностике и прогнозировании в силу их практичности, или обосновывая свой выбор статистически. Можно, например, выбрать равное количество участников в каждой группе с помощью квантилей [90]. С другой стороны, можно получить более полное представление о связи с исходом, выбрав группы с крайними значениями анализируемого показателя, при этом средняя группа (группы) должна быть больше, чем крайние [91]. В исследованиях «случай-контроль» предпочтительна группировка данных исходя из значений показателя в контрольной группе, предназначенной для описания исходной популяции (*source population*). Если точки разделения на группы выбраны из нескольких альтернативных вариантов, но уже после завершения исследования, читателей следует проинформировать об этом. В частности, если точки разделения были выбраны, чтобы минимизировать *p*-значение, истинная сила связи (*strength of an association*) будет преувеличена [81].

При анализе переменных, полученных в результате группирования данных, важно учитывать их исходно непрерывную природу. Например, можно исследовать возможную тенденцию изменения риска в упорядоченных группах. Общий подход — изучать ранги групп как непрерывную переменную. Такая линейность будет приближаться к фактической линейной зависимости, если категории имеют одинаковый интервал (например, возрастные группы с интервалом в 10 лет), но не иначе. D. Il'yasova и соавт. [92] рекомендуют представлять категориальные и непрерывные оценки эффекта со стандартными ошибками, чтобы облегчить метаанализ, равно как и представить ценную информацию о зависимости «доза-ответ». Один анализ может быть источником для другого, и ни один из них не свободен от допущений. Авторы часто игнорируют упорядочение категориальных переменных и рассматривают оценки (и *p*-значения) отдельно для каждой категории в сравнении с референсной категорией. Это может быть уместным при описании, но реальная тенденция к изменению риска по группам может остаться незамеченной. Если такая тенденция обнаружена, доверительный интервал для оценки зависимости одной переменной от другой (*confidence interval for a slope*) будет показателем значимости наблюдения.

12. Статистические методы

12А. Опишите все статистические методы, включая использованные для контроля влияния вмешивающихся факторов

Пример

«Скорректированный относительный риск рассчитывали, используя тест Mantel – Haenszel с целью оценки влияния возраста или пола в сравниваемых группах. Рассчитывали также 95% доверительный интервал скорректированного относительного риска, используя дисперсию, как это было описано в работе S. Greenland и соавт.» [93].

Пояснение

Не существует единственного корректного метода статистического анализа, напротив, есть несколько возможных методов решения одного и того же вопроса, исходя из различных предположений. Поэтому авторы должны предварительно определить тип анализа по крайней мере основных показателей исследования (*primary study objectives*) и зафиксировать его в протоколе. Часто необходим дополнительный анализ вместо или наряду с изначально запланированным. Составляя отчет об исследовании, авторы должны разъяснить читателям, необходимость каких вариантов анализа была определена в ходе проверки данных. Несмотря на то, что различие между запланированным и разведочным (*exploratory*) анализами может быть нечетким, авторам следует разъяснить причины проведения последнего.

Если сравниваемые группы различаются по какому-либо признаку, необходимо проводить коррекцию по потенциально вмешивающимся факторам путем стратификации или с помощью многомерной регрессии (см. вставку 5) [94]. Часто выбор типа регрессионного анализа определяется дизайном исследования. Например, метод регрессии пропорциональных рисков по Коксу (*Cox proportional hazard regression*) обычно используется в когортных исследованиях [95], логистическая регрессия (*logistic regression*) часто становится методом выбора в исследованиях «случай-контроль» [96, 97]. Аналитики должны представить полное описание особенностей процедуры отбора переменных, а не только результаты для окончательной модели [98, 99]. Следует описать результаты сравнения моделей, проведенного для сокращения списка потенциальных вмешивающихся факторов, включаемых в окончательную модель. Даже если одна или две переменные вызывают большую часть очевидных искажений (*confounding*), читатели должны об этом знать. Нестандартные или новые статистические подходы должны быть подкреплены ссылками. Также необходимо указать использованное статистическое программное обеспечение. В качестве руководящего принципа рекомендуем следующее: статистические методы должны быть описаны настолько детально, чтобы квалифицированный читатель при доступе к исходным данным мог проверить результаты исследования [100].

В эмпирическом исследовании только 93 статьи из 169 (55%), представляющие оценки с поправкой на вмешивающиеся факторы, содержали четкую информацию о том, каким образом непрерывные переменные и переменные с множеством категорий были включены в статистическую модель [101]. Другое исследование показало, что среди 67 статей, в которых статистический анализ был выполнен с учетом вмешивающихся факторов, большинство не содержало разъяснений того, как авторы выбирали эти факторы [4].

Вставка 5. Влияние вмешивающихся факторов

Влияние вмешивающихся факторов приводит к искажению эффектов. Может показаться, что исследование либо подтверждает наличие связи между воздействием и риском возникновения заболевания, либо ее опровергает. В действительности кажущаяся связь или ее отсутствие обусловлены другим фактором, который определяет возникновение заболевания и в то же время связан с воздействием. Этот фактор называется «вмешивающимся фактором» или «конфаундером», который приводит к неверной оценке потенциальной «причинно-следственной» связи с воздействием. Например, если женщинам

среднего возраста, у которых повышается артериальное давление, оральные контрацептивы назначаются реже, то простое сравнение частоты сердечно-сосудистых заболеваний между теми, кто принимает контрацептивы, и теми, кто их не принимает, может создать ложное представление о том, что контрацептивы защищают от болезней сердца.

Исследователям следует заранее подумать о потенциальных вмешивающихся факторах. Если предусмотреть их влияние, предварительно изучив всю необходимую информацию, дизайн исследования будет более обоснованным, а собранные данные — полными. Для коррекции влияния вмешивающихся факторов можно использовать ограничение (*restriction*) или подбор пар (*matching*). В приведенном выше примере количество участников исследования может быть ограничено женщинами, не подверженными влиянию вмешивающегося фактора — повышенного артериального давления. Подбор пар по значениям артериального давления также возможен, но нежелателен (см. вставку 2). Чтобы уменьшить влияние вмешивающихся факторов при анализе данных, исследователи могут прибегнуть к стратификации или многомерному анализу. Стратификация заключается в разделении данных на страты (категории) для вмешивающегося фактора (например, страты артериального давления), оценивании связей внутри каждой страты и вычислении общей оценки связи как средневзвешенного значения по всем слоям. Многомерный анализ дает тот же результат, но позволяет учитывать большее количество переменных одновременно. Этот метод более гибкий, но может повлечь за собой ряд дополнительных предположений о математической форме связи между воздействием и заболеванием.

Учет вмешивающихся факторов имеет решающее значение в наблюдательных исследованиях, но у читателей не должно складываться впечатление, что анализ, скорректированный в этом отношении, устанавливает причинно-следственные связи (*association*). Искажение результатов может быть обусловлено еще и остаточным влиянием вмешивающихся факторов (искажение, которое не удалось скорректировать [102]; *residual confounding*), случайной ошибкой выборки (*random sampling error*), систематической ошибкой отбора (*selection bias*) или ошибкой информации (*information bias*) (см. вставку 3).

12Б. Опишите все методы, использованные для изучения подгрупп и взаимодействий

Пример

«Половые различия в восприимчивости к трем факторам риска, связанным с образом жизни, были изучены путем тестирования биологического взаимодействия согласно K.J. Rothman: создали новую композитную переменную с 4 категориями (a^-b^- , a^-b^+ , a^+b^- , a^+b^+) для пола и дихотомической переменной изучаемого воздействия, где a^- и b^- означают отсутствие воздействия. ОР рассчитывали для каждой категории, делая поправку на возраст. Эффект взаимодействия определяли как отклонение от аддитивности абсолютных эффектов и рассчитывали избыточный ОР вследствие взаимодействия (ИОР):

$$\text{ИОР} = \text{ОР}(a^+b^+) - \text{ОР}(a^-b^+) - \text{ОР}(a^+b^-) - 1,$$

где $\text{ОР}(a^+b^+)$ — относительный риск среди тех, кто подвержен воздействию обоих факторов; $\text{ОР}(a^-b^-)$ — референсная категория ($\text{ОР} = 1,0$).

95% доверительный интервал рассчитывали, как предложено D.W. Hosmer и S. Lemeshow. Значение ИОР, равное 0, означало отсутствие взаимодействия» [103].

Пояснение

Как будет подробно обсуждено ниже (см. п. 17), много вопросов вызывают применимость и ценность анализа в ограниченных подгруппах исследуемой популяции [4, 104]. Тем не менее, анализ в подгруппах проводится часто [4]. Читателей необходимо уведомить о том, какие анализы в подгруппах были запланированы заранее, а какие проводились исходя из полученных данных. Также важно разъяснить, какие методы были использованы для выявления различий в величине эффектов или связей между группами (см. п. 17).

О взаимодействии («модификации эффекта») говорят, когда один фактор изменяет эффект другого. Совместное действие 2 факторов можно охарактеризовать двумя способами: используя аддитивную шкалу в терминах разности риска или мультипликативную шкалу — в терминах относительного риска (см. вставку 8). Многие авторы и читатели могут иметь собственные предпочтения в отношении того, как анализировать взаимодействия. Тем не менее, им может быть интересно узнать, в какой степени совместный эффект воздействий отличается от отдельных эффектов. Все согласны с тем, что аддитивная шкала, в которой используются абсолютные риски, больше подходит для общественного здравоохранения и принятия клинических решений [105]. Какой бы ни была точка зрения, она должна быть четко разъяснена читателям, как это сделано в приведенном выше примере [103]. Наиболее информативной может оказаться схема, представляющая отдельные эффекты обоих воздействий и их совместный эффект относительно отсутствия воздействия. Подобная схема приводится в примере п. 17, а расчеты по разным шкалам — во вставке 8.

12В. Объясните, как устраняли последствия отсутствующих (неполных) данных

Пример

«Наш анализ неполных данных основывался на допущении о пропусках данных вследствие случайных причин (*missing at random; MAR*). Для множества многомерных импутаций (подстановки значений) мы использовали метод MICE (*multivariate imputation by chained equations*), реализованный в пакете STATA. Независимо анализировали 10 копий данных, каждая с подстановкой взамен отсутствующих значений, с помощью многомерной логистической регрессии. Мы усреднили полученные оценки переменных для получения одной средней оценки и скорректировали стандартную ошибку в соответствии с правилами D.B. Rubin» [106].

Пояснение

Отсутствующие данные — обычное явление в наблюдательных исследованиях. Разосланные участникам исследования опросники не всегда заполняются целиком, участники пропускают контрольные визиты, источники рутинно собираемых данных и клинические базы данных часто неполны. Несмотря на это, немногие статьи подробно описывают проблему отсутствующих данных [5, 107]. Существуют несколько подходов, которые можно использовать для решения данной проблемы. Их преимущества и недостатки описаны во вставке 6. Мы советуем авторам указать количество отсутствующих данных для каждой из исследуемых переменных (воздействие, исход, вмешивающиеся факторы) на каждом этапе анализа. При описании «потока» участников в ходе исследования авторы должны указать причины отсутствия данных, если это возможно, и указать число участников, исключенных из-за отсутствия данных (см. п. 13). Для анализа, учитываю-

щего отсутствующие данные, следует описать его характер (например, множественные импутации) и сделанные допущения (например, пропуск данных вследствие случайных причин) (см. вставку 6).

Вставка 6. Отсутствующие (неполные) данные: проблемы и способы их решения

Общепринятый подход — ограничить исследование участниками, в отношении которых имеются полные данные по всем переменным для проведения конкретного анализа. Хотя такой анализ «завершенных случаев» зачастую не подвержен систематическим ошибкам, он может быть таковым и всегда будет неэффективным [108]. Систематическая ошибка возникает, если участники с отсутствующими данными нетипичны для всей выборки, а неэффективность — вследствие уменьшения размера выборки.

Использование данных последнего наблюдения для повторных измерений может со временем привести к искажению тенденций, если участники с высокой вероятностью определенного исхода выборочно исключаются из анализа [109]. Добавление показателей отсутствующей категории для вмешивающегося фактора может увеличить остаточное искажение [107]. Метод импутаций, при котором вместо отсутствующего значения указывается предполагаемое или вычисленное значение, может привести к ослаблению или преувеличению исследуемой связи, а без использования сложных методов, описанных ниже, — к слишком малым стандартным ошибкам.

D.B. Rubin разработал типологию проблем, связанных с пропуском данных, основанную на модели вероятности отсутствия наблюдения [108, 110]. Отсутствие данных описывается как полностью случайное (*missing completely at random*; MCAR), если вероятность отсутствия конкретного наблюдения не зависит от значения какой-либо исследуемой переменной или переменных; как случайное (*missing at random*; MAR), если вероятность того, что наблюдения отсутствуют, не зависит от фактических значений отсутствующих данных. К примеру, если предположить, что у детей младшего возраста чаще отмечаются неполные данные спирометрического обследования, вероятность этого не будет связана с истинной ненаблюдаемой функцией легких после поправки к учетом возраста обследуемых. В этом случае отсутствующие значения измерения функции легких будут считаться MAR в статистических моделях, учитывающих возраст. Отсутствие данных может быть неслучайным (*missing not at random*; MNAR), если вероятность этого зависит от пропущенного значения даже после учета имеющихся данных. В случае MNAR для большей убедительности выводы исследования необходимо дополнить разъяснениями того, что привело к потере к данным.

Методы работы с данными категории MAR делятся на три обширных класса [108, 111]: вероятностные подходы [112], взвешенные оценки [113] и множественные импутации [111, 114]. Третий подход — наиболее гибкий и часто используемый, особенно если несколько переменных имеют пропущенные значения [115]. Результаты, полученные с использованием любого из этих подходов, следует сравнить с результатами анализа «завершенных случаев» и обсудить важные различия. Правдоподобность предположений, сделанных при анализе пропущенных данных, как правило, не может быть подтверждена или опровергнута. В частности, невозможно доказать, что пропущенные данные относятся к категории MAR, а не MNAR, поэтому в таких исследованиях рекомендуется проводить анализ чувствительности (*sensitivity analysis*) (см. п. 12Д и 17).

12Г. Когортное исследование. Если применимо, опишите, как решали проблему выбывших из-под наблюдения

Пример

«При проведении лечебных программ с активным наблюдением выявили, что у пациентов, выбывших из-под наблюдения, и тех, кто находился под наблюдением в течение 1 года, исходное количество CD4-клеток было схожим (медиана — 115 и 123 клетки/мкл). Среди пациентов, проходивших программу без активного контроля и потерянных для наблюдения, количество CD4-клеток было значительно ниже, чем среди тех, кто оставался под наблюдением (медиана — 64 и 123 клетки/мкл). <...> Данные лечебной программы с пассивным контролем результатов были исключены из последующего анализа» [116].

Пояснение

При анализе данных когортных исследований используют таблицы смертности или иные подходы, основанные на человеко-времени наблюдения и времени до развития изучаемого заболевания. Предполагается, что среди лиц, у которых по окончании периода наблюдения не отмечали признаков заболевания, время наблюдения не связано с вероятностью развития исхода. Такое предположение будет верным, если период наблюдения заканчивается определенной датой или по достижении определенного возраста. Если участники прекращают участие в исследовании до наступления этой даты, речь идет о выбывании из-под наблюдения (*loss to follow-up*). Достоверность (*validity*) исследования снижается, если выборочно выбывают участники, подвергшиеся воздействию, или пациенты с более высоким риском развития заболевания (цензурирование информации; *informative censoring*). В приведенном выше примере выбывавшие из-под наблюдения пациенты, получавшие лечение по программе, не предусматривавшей активного контроля, имели меньше CD4-хелперов, чем те, за кем продолжали наблюдать, и, следовательно, имели более высокий риск летального исхода [116].

Важно проводить различие между пациентами, завершившими исследование и выбывшими из него. К сожалению, статистические программы обычно не различают эти ситуации и в обоих случаях автоматически обрывают («цензурируют») время отслеживания исходов в момент окончания наблюдения. Поэтому исследователям необходимо решить, желательно еще на этапе планирования исследования, как поступить с данными о пациентах, выбывших из-под наблюдения. Если потери невелики, исследователи могут исключить из анализа данные о пациентах, не завершивших наблюдение, или обрабатывать эти данные, предполагая, что участники умерли или к моменту выбывания из-под наблюдения, или к окончанию исследования. Советуем авторам указывать, сколько пациентов выбыло из-под наблюдения и какая стратегия цензурирования данных была использована.

12Г. Исследование «случай-контроль».

Если применимо, объясните, как проводили подбор (*matching*) «случаев» и «контролей»

Пример

«Для сравнения пациентов с деменцией и подобранных им в пары «контролей» с целью оценки факторов риска сердечно-сосудистых заболеваний, развития спонтанной эмболии сосудов мозга, поражения сонных артерий и артериовенозных шунтов использовали тест McNemar, парный *t*-тест и условную логистическую регрессию» [117].

Пояснение

В исследованиях «случай-контроль» с индивидуальным подбором пар определение отношения шансов без учета факта подбора обычно приводит к смещению оценок к единице (см. вставку 2). Поэтому зачастую необходим парный анализ (*matched analysis*). Интуитивно это можно представить как стратифицированный анализ, где каждый случай рассматривается как одна страта для своих «контролей». Анализ основывается на предположении, что «случаи» подвергаются воздействию чаще, чем «контроли», несмотря на то, что они одинаковы по признаку, по которому проводится подбор. Исследователи могут проводить такой стратифицированный анализ, используя метод Mantel – Haenszel для таблиц 2×2 . В его простейшем варианте отношение шансов становится отношением пар, дискордантных по оказываемому на них воздействию. Если подбор происходит по таким универсальным характеристикам, как пол или возраст, нет необходимости в индивидуальном («пациент к пациенту») анализе, достаточно простого анализа по категориям пола и возраста [50]. Для других переменных, используемых для подбора пар, таких как соседство, родство или дружба, каждый подобранный набор «случаев» и «контролей» необходимо рассматривать как отдельную страту.

В исследованиях с индивидуальным подбором пар наиболее широко используется метод условной логистической регрессии: каждый «случай» и его «контроли» рассматриваются вместе. Этот метод необходим, когда число «контролей» для разных «случаев» неодинаково и когда кроме переменных, используемых для подбора пар, есть и другие переменные, которые необходимо использовать для коррекции. Для того чтобы читатели могли решить, был ли дизайн исследования с подобранными парами (*matched design*) принят в расчет при анализе, рекомендуем авторам детально описывать примененные статистические методы анализа данных. Если учет подбора пар мало влияет на полученные оценки, авторы могут прибегнуть к анализу, используемому для сравнения независимых выборок.

12Г. Одномоментное исследование.

Если применимо, опишите аналитические методы, учитывающие стратегию формирования выборки

Пример

«Стандартную ошибку рассчитывали, используя метод разложения Тейлора для оценки выборочной ошибки при комплексном дизайне выборки. <...> Выявили, что общее влияние дизайна исследования для диастолического артериального давления составило 1,9 у мужчин и 1,8 у женщин, для систолического артериального давления — 1,9 и 2,0 соответственно» [118].

Пояснение

В большинстве одномоментных исследований используется предварительно описанная стратегия формирования выборки (*sampling strategy*) из исходной популяции (*source population*). Стратегия может быть более сложной, чем простой случайный отбор. Она может включать несколько этапов, в том числе кластеризацию участников (например, по районам или деревням). Пропорциональная стратификация может гарантировать правильное представление подгрупп с определенной характеристикой. Непропорциональная стратификация может быть полезной для дополнительной выборки подгруппы, представляющей особый интерес.

Оценка связей в комплексной выборке может быть более или менее точной, чем в простой случайной выборке. Критерии точности, такие как стандартная ошибка или доверительный интервал, должны быть представлены с учетом эффекта дизайна (*design effect*), т.е. показателя отношения, который описывает, насколько точным будет исследование, если используется более сложная стратегия вместо простого случайного отбора [119]. Большинство сложных методик формирования выборки исследования приводят к снижению точности, в результате чего эффект дизайна превышает единицу.

Советуем авторам четко описать методы, используемые для коррекции данных в комплексных стратегиях формирования выборки, что поможет читателям понять, как выбор метода создания выборки влияет на точность получаемых оценок. Например, при кластерной выборке неявный компромисс между более легким методом сбора данных и потерей точности очевиден, если авторы сообщают о влиянии дизайна. В приведенном выше примере расчеты показывают, что влияние дизайна составило 1,9 для мужчин. Это демонстрирует, что необходимый размер выборки должен быть в 1,9 раза больше, чем при простом случайном отборе для получения оценок аналогичной точности.

12Д. Опишите использованные анализы чувствительности

Пример

«Поскольку в исследовании оказалась высокой доля отсутствующих данных по причине смерти участников (38 из 148; 25,7%) по сравнению с выжившими (15 из 437; 3,4%) <...>, это могло привести к возникновению систематической ошибки. Поэтому мы провели анализ чувствительности. Предположили, что доля женщин, принимающих оральные контрацептивы (19,1% среди умерших, 11,4% среди выживших), в исследуемой группе может быть применима к популяции в целом. Далее учитывали два крайних сценария: либо все подвергнутые воздействию и выбывшие из-под наблюдения принимали контрацептивы второго поколения, либо все — третьего» [120].

Пояснение

Анализ чувствительности или устойчивости результатов исследования к изменениям исходных условий (*sensitivity analyses*) помогает исследователям понять, согласуются ли основные результаты исследования с теми, что могли бы быть получены при использовании альтернативных подходов к анализу данных или допущений [121]. Таким образом можно оценить влияние на результат критериев включения в анализ, вариантов определений воздействия и исхода [122], вмешивающихся факторов, отсутствующих данных [120, 123], возможных систематических ошибок отбора, неточного и неодинакового определения воздействия, заболеваний и других переменных, выбора методов анализа, например обработки количественных переменных (см. п. 11). Для создания моделей одновременного влияния нескольких систематических ошибок или допущений все чаще используются более сложные методы [124–126].

В 1959 г. J. Cornfield и соавт. продемонстрировали, что относительный риск возникновения рака легких у курильщиков по сравнению с некурящими равен 9, и совершенно невероятно, чтобы это могло объясняться другими вмешивающимися факторами, поскольку таковые должны были бы встречаться у курильщиков минимум в 9 раз чаще, чем среди некурящих [127]. Этот анализ не исклю-

чил возможности наличия такого фактора, но определил, какой должна быть распространенность такого фактора. Аналогичный подход недавно был использован для выявления правдоподобных вмешивающихся факторов, объясняющих связь между возникновением лейкемии у детей и проживанием вблизи линий электропередач [128]. В общем анализ чувствительности может использоваться для определения степени влияния вмешивающихся факторов (*degree of confounding*), систематических ошибок отбора (*selection bias*) или ошибок информации (*information bias*). Важным и, возможно, недостаточно признанным является использование анализа чувствительности в случае, когда исследование демонстрирует незначительную связь между воздействием и исходом либо ее отсутствие, с целью выявления действия вмешивающихся факторов или систематических ошибок, смещающих результаты в сторону нуля.

Результаты

Раздел «Результаты» должен содержать описание обнаруженных фактов: от набора участников исследования, описания выборки исследования до основных результатов и дополнительных анализов. Здесь не должно быть интерпретаций и рассуждений, отражающих взгляды и мнения авторов.

13. Участники

13А. Укажите количество лиц на каждом этапе исследования: число потенциально подходящих для участия в исследовании, обследованных на предмет соответствия критериям отбора, допущенных к участию, включенных в исследование, завершивших последующее наблюдение, а также участников, чьи результаты были проанализированы

Пример

«Из выборки 105 отдельно расположенных баров и таверн 13 заведений больше не работали, 9 располагались при ресторанах, оставшиеся 83 отвечали условиям отбора. Собственникам 22 баров не удалось дозвониться после 6 и более попыток. Собственники 36 баров отказались участвовать в исследовании. <...> В 25 барах и тавернах, принявших участие в исследовании, состояли в штате 124 бармена, 67 из них как минимум 1 раз в неделю работали в дневную смену. Из числа последних 54 бармена (81%) были полностью проинтервьюированы в начале исследования и прошли спирометрию, 53 из них (98%) завершили период наблюдения» [129].

Пояснение

Детальная информация о том, как происходил набор участников в исследование, важна по нескольким причинам. Включенные в исследование часто по разным причинам отличаются от целевой популяции (*target population*), к которой применяются результаты исследования. Это может привести к оценкам распространенности (*prevalence*) и инцидентности (*incidence*), которые не отражают положение дел в целевой популяции. Например, лица, соглашающиеся участвовать в почтовом опросе об особенностях сексуального поведения, посещают церковь реже, придерживаются менее консервативных взглядов на секс и получили первый сексуальный опыт в более раннем возрасте, а также с большей вероятностью курят сигареты и употребляют алкоголь, чем лица, отказавшиеся участвовать в таком исследовании [130]. Эти различия позволяют предположить, что в ходе почтового опроса можно переоценить сексуальный либе-

рализм и активность популяции. Такая систематическая ошибка отклика (*response bias*) (см. вставку 3) может исказить связь воздействия и заболевания, если эта связь различается у подходящих для включения в исследование и включенных в него. Другой пример: связь между ранним материнством и возникновением лейкемии у потомства, продемонстрированная в нескольких исследованиях «случай-контроль» [131, 132], объяснялась несопоставимостью молодых женщин в группах «случаи» и «контроли». Юные матери здоровых детей были менее склонны участвовать в исследованиях, чем те, чьи дети болели [133]. Хотя низкий уровень участия не обязательно ставит под сомнение достоверность (*validity*) исследования, прозрачность информации об участии и причинах неучастия имеет важное значение. Кроме того, поскольку не существует универсального определения таких понятий, как «доля приглашенных участвовать» (*participation rate*), «доля согласившихся участвовать» (*response rate*), «доля завершивших наблюдение» (*follow-up rate*), необходимо пояснять читателям, как авторы рассчитывали эти показатели [134].

В идеале исследователи должны проводить подсчет числа индивидуумов на каждом этапе набора участников: от определения целевой популяции до включения данных об участнике в анализ. В зависимости от типа исследования может потребоваться подсчет числа потенциально подходящих участников (*potentially eligible*), проверенных на соответствие критериям отбора (*assessed for eligibility*), признанных подходящими для участия (*to be eligible*), включенных в исследование (*included in the study*), прошедших обследование (*examined*), продолживших участие (*followed up*) и включенных в анализ (*included in the analysis*). Если выборка участников исследования формируется в 2 и более стадии (многоступенчатый отбор), как это произошло в описанном выше примере, может потребоваться информация о составе выборки на каждом из этапов. В исследованиях «случай-контроль» мы советуем авторам описывать «поток» (последовательность и результаты формирования выборки) участников отдельно для групп «случаев» и «контролей» [135]. Иногда «контроли» отбираются из нескольких источников, например, госпитализированные пациенты и обитатели домов престарелых. В этом случае рекомендуем отдельно подсчитывать число участников каждого типа. S.H. Olson и соавт. разработали полезные рекомендации по составлению отчетов о подборе «контролей» путем случайного обзвона и применения других методов [136].

Недавний обзор эпидемиологических исследований, опубликованный в 10 журналах по общей эпидемиологии, общественному здравоохранению и медицине, показал, что некоторая информация об участии была предоставлена в 47 исследованиях «случай-контроль» из 107 (59%), 49 когортных исследованиях из 154 (32%) и 51 одномерном исследовании из 86 (59%) [137]. В двух других обзорах литературы было также показано, что публикуемые сведения об участии или неучастии в эпидемиологических исследованиях были неполными либо отсутствовали вовсе [4, 5]. Наконец, есть свидетельства того, что участие в эпидемиологических исследованиях могло снизиться в последние десятилетия [137, 138], что свидетельствует о необходимости прозрачной отчетности [139].

13Б. Укажите причины выбывания на каждом этапе исследования

Пример

«Основными причинами невключения были тяжелая болезнь или смерть потенциального участника до про-

ведения интервью (30% в группе „случаев“, менее 1% в группе „контролей“), неполучение ответа (2% в группе „случаев“, 21% в группе „контролей“), отказ от участия (10% в группе „случаев“, 29% в группе „контролей“) и другие причины (отказ от сотрудничества с консультантом или врачом общей практики, незнание английского языка, психические нарушения) (7% в группе „случаев“, 5% в группе „контролей“) [140].

Пояснение

Раскрытие причин невключения, а также прекращения участия в исследовании или исключения информации об участниках из статистического анализа помогает читателям прийти к заключению, репрезентативна ли выборка исследования (*study population*) целевой популяции (*target population*) и был ли риск возникновения систематической ошибки. Например, в опросе о здоровье неучастие в силу причин, не относящихся к состоянию здоровья (например, неполучение письма-приглашения из-за неверно указанного адреса), может влиять на точность оценок, но почти наверняка не приведет к систематической ошибке. В противоположность этому, если многие уклонятся от участия в опросе из-за болезни или, наоборот, из-за ощущения себя здоровыми, это может привести к переоценке или недооценке распространенности проблем со здоровьем в популяции.

13В. Рассмотрите возможность использования потоковой диаграммы

Пример

См. рисунок.

Пояснение

Информативная и хорошо структурированная потоковая диаграмма (как в примере выше) помогает сделать информацию более доступной и легко читаемой, чем подробное описание в тексте [142]. Диаграмма может включать основные результаты, такие как число основных исходов. Хотя мы рекомендуем использовать потоковую диаграмму, в частности, для сложных наблюдательных исследований, мы не предлагаем для нее конкретный формат.

14. Описательные данные

14А. Охарактеризуйте участников исследования (например, укажите демографические, клинические, социальные характеристики).

Приведите данные о воздействиях и потенциальных вмешивающихся факторах

Пример

См. табл. 2.

Пояснение

Описание участников и изучаемого воздействия необходимо читателям для того, чтобы судить об обобщаемости (*generalisability*) результатов. Информация о потенциальных вмешивающихся факторах, включающая описание способов их измерения, влияет на суждение о достоверности исследования (*study validity*). Мы советуем авторам описывать непрерывные переменные для каждой группы исследования, указывая среднее арифметическое и стандартное отклонение или, в случае асимметричного распределения, медиану и перцентильный разброс (например, 25-й и 75-й перцентили). Переменные, включающие небольшое число упорядоченных категорий (например,

стадии заболевания от I до IV), не должны быть представлены в виде непрерывных переменных, лучше указывать число и долю для каждой категории (см. вставку 4). В исследованиях с группами сравнения описательные характеристики и числа должны приводиться для каждой группы, как в примере выше.

Расчетные величины, такие как стандартная ошибка и доверительный интервал, не должны использоваться для описания варибельности характеристик, а в описательных таблицах не следует приводить результаты статистических тестов. Кроме того, *p*-значения не подходят для отбора вмешивающихся факторов с целью коррекции анализа; даже небольшое различие групп в распределении значений фактора, который сильно влияет на исход, может быть важным [144, 145].

В когортных исследованиях может оказаться полезным документировать, как воздействие влияло на другие характеристики и вмешивающиеся факторы. Авторы могут представить эту информацию в таблице со столбцами для двух или более категорий воздействия, что позволит судить о разнице вмешивающихся факторов между этими категориями.

В исследованиях «случай-контроль» суждение о потенциальных вмешивающихся факторах не может выноситься на основании сравнения «случаев» и «контролей». Лица, включенные в контрольную группу, представляют исходную популяцию (*source population*) и обычно отличаются от группы «случаев» по многим аспектам. Например, в исследовании, посвященном оральным контрацептивам и инфаркту миокарда, молодые женщины с инфарктом миокарда отличались более высокой концентрацией холестерина, курением, отягощенным семейным анамнезом, чем участницы группы «контролей» [146]. Это не влияло на оценку воздействия оральных контрацептивов, в том числе потому, что при прописывании этих лекарственных средств не руководствовались наличием факторов риска, поскольку их оценку проводили после развития инфаркта (см. вставку 5). В исследованиях «случай-контроль» эквивалент сравнения подвергшихся и не подвергавшихся воздействию на наличие вмешивающихся факторов (как это делается в когортных исследованиях) может быть достигнут путем изучения исходной для случаев популяции: если контрольная группа достаточно велика и представляет исходную популяцию, подвергшиеся и не подвергавшиеся воздействию «контроли» можно сравнить на наличие потенциальных вмешивающихся факторов [121, 147].

14Б. Для каждой из анализируемых переменных укажите количество участников с пропущенными данными

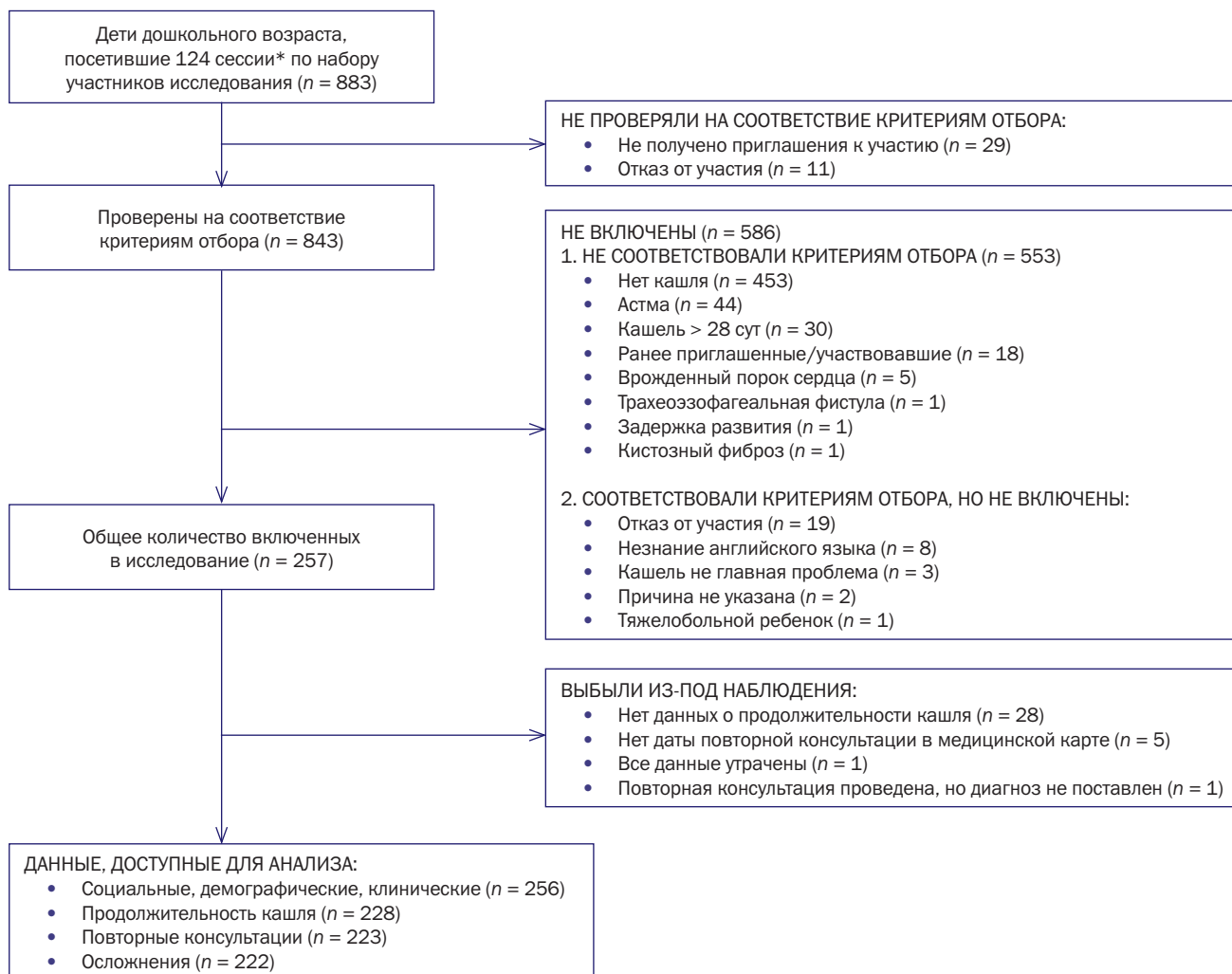
Пример

См. табл. 3.

Пояснение

Поскольку отсутствующие данные могут быть источником систематической ошибки или влиять на обобщаемость результатов, авторы должны сообщить читателям о количестве таких данных применительно к воздействию, потенциальным вмешивающимся факторам и другим важным характеристикам пациентов (см. п. 12В и вставку 6). В когортном исследовании необходимо указать количество участников, выбывших из-под наблюдения, и причины выбывания, поскольку незавершенная процедура наблюдения может исказить результаты (см. п. 12Г и 13) [148]. Мы советуем авторам использовать таблицы и рисунки для представления этой информации.

Рисунок. Поток диаграмма (адаптировано из [141])
Figure. Flow diagram (adapted from [141])



Примечание. <*> — отсутствуют данные об одной из сессий по набору участников, на которой присутствовали не менее 3 посетителя с кашлем, отобрано 2 участника.

Note. Denominator data missing from one session at which at least 3 attended with cough, 2 recruited.

Таблица 2. Характеристики участников при включении в исследование, Castellana G (Италия), 1985–1986 гг. (адаптировано из [143])
Table 2. Characteristics of the Study Base at Enrolment, Castellana G (Italy), 1985–1986 (adapted from [143])

	НСV-отрицательные (n = 1458)	НСV-положительные (n = 511)	Неизвестно (n = 513)
Пол (%)			
Мужской	936 (64%)	296 (58%)	197 (39%)
Женский	522 (36%)	215 (42%)	306 (61%)
Средний возраст при регистрации (СО)	45,7 (10,5)	52,0 (9,7)	52,5 (9,8)
Ежедневное потребление алкоголя (%)			
Нет	250 (17%)	129 (25%)	119 (24%)
Умеренное *	853 (59%)	272 (53%)	293 (58%)
Чрезмерное**	355 (24%)	110 (22%)	91 (18%)

Примечание. СО — стандартное отклонение. НСV — вирус гепатита С. <*> — мужчины < 60 г этанола/сут, женщины < 30 г этанола/сут; <***> — мужчины > 60 г этанола/сут, женщины > 30 г этанола/сут.

Note. SD (CO) — standard deviation. HCV — hepatitis C virus. <*> — males < 60 g ethanol/day, females < 30 g ethanol/day; <***> — males > 60 g ethanol/day, females > 30 g ethanol/day.

Таблица 3. Симптомы заболевания как конечные точки в анализе выживаемости (адаптировано из [141])**Table 3.** Symptom End Points Used in Survival Analysis (adapted from [141])

	Кашель	Одышка	Бессонница
Симптом устранен	201 (79%)	138 (54%)	171 (67%)
Цензурирование	27 (10%)	21 (8%)	24 (9%)
Симптомы отсутствовали	0	46 (18%)	11 (4%)
Данные отсутствуют	28 (11%)	51 (20%)	50 (20%)
Всего	256 (100%)	256 (100%)	256 (100%)

14В. Когортное исследование. Опишите характеристики времени наблюдения (например, укажите среднюю и общую продолжительность наблюдения)

Пример

«На протяжении 4366 человеко-лет наблюдения (медиана длительности наблюдения — 5,4 года, максимум — 8,3 года) у 265 участников была диагностирована деменция, в том числе у 202 — болезнь Альцгеймера» [149].

Пояснение

Читателям необходимо знать длительность и границы периода наблюдения за участниками, в течение которого была получена информация об исходах. Авторы могут представить сводку этих данных, указав значения среднего арифметического и/или медианы времени наблюдения. Умножение средней продолжительности наблюдения на число участников исследования позволит читателям рассчитать общее количество человеко-лет. Авторы могут также привести минимальное и максимальное время наблюдения или перцентили распределения. Можно указать общую длительность наблюдения в человеко-годах или долю данных, которые были собраны [148]. Вся информация такого рода может быть представлена отдельно для участников в соответствии с 2 и более категориями воздействия. Почти половина из 132 статей в журналах по онкологии (в основном когортные исследования) не содержала описания данных о продолжительности наблюдения [37].

15. Исходы

Когортное исследование. Укажите количество событий исхода или представьте сводку основных показателей за время наблюдения

Пример

См. табл. 4.

Исследование «случай-контроль». Укажите количество событий по каждой категории воздействия или сводные показатели воздействия

Пример

См. табл. 5.

Одномоментное исследование. Укажите количество событий исхода или описание основных показателей исследования

Пример

См. табл. 6.

Пояснение

Прежде чем обратиться к возможной связи между воздействиями (факторами риска) и исходами, авторы должны привести некоторые относящиеся к теме описательные данные. Оценку такой связи можно представить в той же таблице, в которой представлены описательные данные (см. п. 14А). В когортном исследовании при изучении событий в качестве исходов (*outcomes*) укажите количество событий по каждому анализируемому исходу.

Таблица 4. Случаи сероконверсии ВИЧ-1 по отдельным социально-демографическим переменным: 1990–1993 гг. (адаптировано из [150])**Table 4.** Rates of HIV-1 Seroconversion by Selected Sociodemographic Variables: 1990–1993 (adapted from [150])

Переменные	Человеко-годы	Количество участников с сероконверсией	Число событий / 1000 человеко-лет (95% ДИ)
Календарный год			
1990	2197,5	18	8,2 (4,4–12,0)
1991	3210,7	22	6,9 (4,0–9,7)
1992	3162,6	18	5,7 (3,1–8,3)
1993	2912,9	26	8,9 (5,5–12,4)
1994	1104,5	5	4,5 (0,6–8,5)
Племена			
Баганда	8433,1	48	5,7 (4,1–7,3)
Другие народы Уганды	578,4	9	15,6 (5,4–25,7)
Руандийцы	2318,6	16	6,9 (3,5–10,3)
Другие племена	866,0	12	13,9 (6,0–21,7)
Религия			
Мусульмане	3313,5	9	2,7 (0,9–4,5)
Другая религия	8882,7	76	8,6 (6,6–10,5)

Примечание. ДИ — доверительный интервал.

Note. CI (ДИ) — confidence interval.

Таблица 5. Факторы экспозиции среди «случаев» и «контролей» с циррозом печени (адаптировано из [151])**Table 5.** Exposure among Liver Cirrhosis Cases and Controls (adapted from [151])

	«Случаи» (n = 40)	«Контроли» (n = 139)
Винилхлоридный мономер (совокупное воздействие: ч/млн × годы)		
< 160	7 (18%)	38 (27%)
160–500	7 (18%)	40 (29%)
500–2500	9 (23%)	37 (27%)
> 2500	17 (43%)	24 (17%)
Потребление алкоголя (г/день)		
< 30	1 (3%)	82 (59%)
30–60	7 (18%)	46 (33%)
> 60	32 (80%)	11 (8%)
HBsAG/HCV		
Отрицательный	33 (83%)	136 (98%)
Положительный	7 (18%)	3 (2%)

Примечание. HBsAG — поверхностный антиген вируса гепатита В; HCV — вирус гепатита С.

Note. HBsAG — hepatitis B surface antigen; HCV — hepatitis C virus.

Таблица 6. Распространенность астмы и диагностированной сенной лихорадки в зависимости от среднего уровня антигена плесневого грибка *Alternaria alternata* в жилых помещениях (адаптировано из [152])**Table 6.** Prevalence of Current Asthma and Diagnosed Hay Fever by Average *Alternaria alternata* Antigen Level in the Household (adapted from [152])

Категории содержания грибка <i>Alternaria</i> *	Астма		Сенная лихорадка	
	N	Распространенность** (95% ДИ)	N	Распространенность** (95% ДИ)
1-й тертиль	40	4,8 (3,3–6,9)	93	16,4 (13,0–20,5)
2-й тертиль	61	7,5 (5,2–10,6)	122	17,1 (12,8–22,5)
3-й тертиль	33	8,7 (6,7–11,3)	93	15,2 (12,1–18,9)

Примечание. <*> — 1-й тертиль < 3,90 мкг/г; 2-й тертиль от 3,90 до 6,27 мкг/г; 3-й тертиль ≥ 6,28 мкг/г. <***> — проценты (95% ДИ) взвешены для многоступенчатого дизайна формирования выборки Национального исследования свинца и аллергенов в жилых помещениях. ДИ — доверительный интервал.

Note. <*> — 1st tertile < 3.90 µg/g; 2nd tertile 3.90–6.27 µg/g; 3rd tertile ≥ 6.28 µg/g. <***> — percentage (95% CI) weighted for the multistage sampling design of the National Survey of Lead and Allergens in Housing. CI (ДИ) — confidence interval.

При необходимости опишите частоту событий в пересчете на человеко-год наблюдений (*event rate per person-year*). Если риск событий изменялся в период наблюдения, представьте число и частоту событий, разбив период наблюдения на подходящие интервалы с помощью таблицы дожития (смертности) или графика Kaplan – Meier. Может оказаться предпочтительным использовать график кумулятивной инцидентности (*cumulative incidence*), который возрастает от 0% быстрее, чем замедляется при приближении к 100%, особенно если частота наблюдаемых событий составляет менее 30% [153]. Рассмотрите возможность представления такой информации отдельно для участников в различных категориях воздействия. Если в когортном исследовании изучают и другие зависящие от времени исходы (например, количественные маркеры заболевания, такие как артериальное давление), укажите, как изменялись соответствующие сводные оценки (например, среднее арифметическое и стандартное отклонение) в течение времени, возможно — в виде таблицы или рисунка.

Для одномоментных исследований мы рекомендуем представлять, как описано выше, информацию о распространенности событий исходов или сводные оценки измере-

ний. В исследованиях «случай-контроль» основное внимание должно быть уделено описанию частотных или количественных характеристик воздействий отдельно для «случаев» и «контролей» [154]. При любом дизайне исследования может быть полезно представить непрерывные данные исходов или воздействий в таблице с разбивкой по категориям, даже если эти данные не анализируются как таковые.

16. Основные результаты

16А. Укажите нескорректированные оценки и, если применимо, оценки с поправкой на вмешивающиеся факторы и их точность (например, 95% ДИ). Объясните, влияние каких вмешивающихся факторов было учтено и почему именно эти факторы были выбраны

Пример 1

«Изначально мы рассматривали включение следующих переменных в качестве возможных вмешивающихся факторов в стратифицированный анализ Mantel – Haenszel. <...> В окончательную логистическую регрессионную модель включили следующие переменные <...>, что вызвало 10% изменение отношения шансов после поправки по Mantel – Haenszel» [155].

Пример 2
См. табл. 7.

Пояснение

Во многих случаях авторы могут представлять читателям результаты нескорректированного, минимально скорректированного и полностью скорректированного анализа. Мы советуем приводить результаты нескорректированного анализа вместе с основными данными, например, указывать число «случаев» и «контролей», подвергшихся и не подвергавшихся воздействию. Это позволит читателям разобраться в данных, прежде чем переходить к анализу связей (см. также п. 15). Для скорректированных анализов укажите число лиц, включенных в анализ, поскольку это число может отличаться от такового при нескорректированном анализе вследствие неполных данных у некоторых переменных, используемых для коррекции (см. п. 12В). Оценки необходимо приводить вместе с доверительными интервалами.

Читатели могут сравнивать нескорректированные показатели связи с теми, которые скорректированы с учетом потенциальных вмешивающихся факторов, и судить, насколько и в каком направлении они изменились. Читатели могут подумать, что «скорректированные» результаты представляют собой «причинную» часть показателей связи, однако скорректированные результаты не обязательно свободны от случайной ошибки выборки (*random sampling error*), систематических ошибок отбора (*selection bias*) и информации (*information bias*), остаточного влияния вмешивающихся факторов (*residual confounding*) (см. вставку 5). Поэтому интерпретировать скорректированные результаты надо с большой осторожностью, их достоверность часто зависит от полноты информации о важных вмешивающихся факторах, точности их измерения и выбора подходящей статистической модели (см. п. 20) [157, 158].

Авторы должны объяснить все рассматриваемые ими потенциальные вмешивающиеся факторы, а также критерии включения или исключения переменных в статистические модели. При принятии решения о включении или исключении переменных нужно руководствоваться знаниями, четко сформулированными предположениями или причинными отношениями. Неверные решения могут привести к систематическим ошибкам, например, при включении переменных, вовлеченных в причинную цепочку между воздействием и заболеванием (кроме случаев, когда цель заключается в оценке того, какой эффект оказывает промежуточная переменная). Если решение о включении переменной в модель основано на изменении оценки, важно указать, какие изменения было решено считать достаточно значимыми для включения переменной в анализ. Если при отборе вмешивающихся факторов использовали стратегии «обратного исключения» (*“backward deletion”*) или «прямого включения» (*“forward inclusion”*), этот процесс должен быть разъяснен и указаны уровни значимости для отклонения нулевой гипотезы об отсутствии влияния вмешивающихся факторов. Следует отметить, что мы не советуем выбирать вмешивающиеся факторы, основываясь исключительно на проверке статистической значимости [147, 159, 160].

Согласно недавним работам, посвященным оценке качества отчетов эпидемиологических исследований, доверительные интервалы были указаны в большинстве статей [4]. Однако лишь немногие авторы объяснили свой выбор вмешивающихся факторов [4, 5].

16Б. При категоризации непрерывных переменных укажите границы категорий

Пример
См. табл. 8.

Таблица 7. Отношение частот повторной госпитализации при лечении пациентов, находящихся в стационаре по месту жительства, после первой госпитализации в связи с шизофренией и шизоаффективными расстройствами (адаптировано из [156])

Table 7. Relative Rates of Rehospitalization by Treatment in Patients in Community Care after First Hospitalization due to Schizophrenia and Schizoaffective Disorder (adapted from [156])

Лечение	Количество рецидивов	Человеко-годы	Отношение частот (95% ДИ)	Скорректированное отношение частот (95% ДИ)	Полностью скорректированное отношение частот (95% ДИ)
Перфеназин	53	187	0,41 (0,29–0,59)	0,45 (0,32–0,65)	0,32 (0,22–0,49)
Оланзапин	329	822	0,59 (0,45–0,75)	0,55 (0,43–0,72)	0,54 (0,41–0,71)
Клозапин	336	804	0,61 (0,47–0,79)	0,53 (0,41–0,69)	0,64 (0,48–0,85)
Хлорпротиксен	79	146	0,79 (0,58–1,09)	0,83 (0,61–1,15)	0,64 (0,45–0,91)
Тиоридазин	115	201	0,84 (0,63–1,12)	0,82 (0,61–1,10)	0,70 (0,51–0,96)
Перфеназин	155	327	0,69 (0,58–0,82)	0,78 (0,59–1,03)	0,85 (0,63–1,13)
Рisperидон	343	651	0,77 (0,60–0,99)	0,80 (0,62–1,03)	0,89 (0,69–1,16)
Галоперидол	73	107	1,00	1,00	1,00
Хлорпромазин	82	127	0,94 (0,69–1,29)	0,97 (0,71–1,33)	1,06 (0,76–1,47)
Левомепромазин	52	63	1,21 (0,84–1,73)	0,82 (0,58–1,18)	1,09 (0,76–1,57)
Без антипсихотической терапии	2248	3362	0,98 (0,77–1,23)	1,01 (0,80–1,27)	1,16 (0,91–1,47)

Примечание. Скорректировано с учетом пола, календарного года, возраста на момент начала наблюдения, количества предыдущих рецидивов, продолжительности первой госпитализации и последующего наблюдения (скорректированные показатели), а также дополнительно по показателю предрасположенности (*score of the propensity*) к началу лечения препаратами, отличными от галоперидола (полностью скорректированные показатели). ДИ — доверительный интервал.

Note. CI (ДИ) — confidence interval. Adjusted for sex, calendar year, age at onset of follow-up, number of previous relapses, duration of first hospitalization, and length of follow-up (adjusted column) and additionally for a score of the propensity to start a treatment other than haloperidol (fully adjusted column). CI (ДИ) — confidence interval.

Таблица 8. Полихлорированные бифенилы в сыворотке пуповинной крови (адаптировано из [161])**Table 8.** Polychlorinated Biphenyls in Cord Serum (adapted from [161])

Квартиль	Диапазон, нг/г	Количество
1	0,07–0,24	180
2	0,24–0,38	181
3	0,38–0,60	181
4	0,61–18,14	180

Пояснение

Категоризация значений непрерывной переменной имеет несколько важных последствий для анализа (см. вставку 4) и влияет на представление результатов. В таблицах исходы должны быть приведены для каждой категории фактора воздействия (*exposure category*), например, с указанием числа лиц, подвергшихся воздействию (*persons at risk*), человеко-времени для периода воздействия (*person-time at risk*), если применимо — то отдельно для каждой группы (например, для «случаев» и «контролей»). Детальное описание категорий может помочь при сравнении исследований и проведении мета-анализа. Если данные группировали с использованием традиционных точек разделения (*cut-points*), таких как пороговые значения индекса массы тела [162], границы групп (т.е. диапазон значений) легко представить, за исключением самой верхней и нижней категории. Если разделение на категории выполнено по значениям квантилей, границы групп не могут быть выведены из данных. Поэтому авторы как минимум должны указать границы категории; полезно также указать диапазон данных, среднее арифметическое значение или медиану для значений показателя внутри категорий.

16В. Если применимо, рассмотрите возможность преобразования относительного риска в абсолютный для значимого периода времени

Пример

«10-летнее использование гормонально-заместительной терапии привело к 5 (95% ДИ 3–7) дополнительным случаям рака молочной железы на 1000 женщин, получавших только эстрогенные препараты, и 19 (95% ДИ 15–23) дополнительным случаям рака на 1000 женщин, получавших комбинацию эстрогенов и прогестинов» [163].

Пояснение

Результаты исследований, в которых изучалась связь (*association*) между воздействием и заболеванием, часто представляют в виде относительных величин: отношение рисков, частот или шансов (см. вставку 8). Относительные показатели отражают силу такой связи. Если относительный риск (*relative risk*) сильно отличается от 1, маловероятно, что связь обусловлена вмешивающимися факторами [164, 165]. Относительные эффекты или ассоциации, как правило, более согласованы между исследованиями и популяциями, чем абсолютные показатели, но то, что часто имеет место, может быть неуместным в конкретном случае. Например, схожий относительный риск был выявлен при изучении классических факторов риска сердечно-сосудистых заболеваний у мужчин, живущих в Северной Ирландии, Франции, Соединенных Штатах Америки и в Германии, хотя базовый уровень риска ишемической болезни сердца различается между этими странами весь-

ма значительно [166, 167]. В противоположность этому в исследовании артериальной гипертензии как фактора риска смерти от сердечно-сосудистых заболеваний данные были более согласованы с постоянной (во времени) разницей частот (*constant rate difference*), чем с постоянным отношением частот (*constant rate ratio*) [168].

Широко используемые статистические модели, в том числе логистической регрессии [169] и модели пропорциональных рисков Кокса [170], основываются на относительных величинах. В этих моделях легко выявить лишь отклонения относительных характеристик эффекта от постоянного уровня. Тем не менее, показатели отклонения от аддитивности разницы рисков, такие как избыточный относительный риск вследствие взаимодействия (*relative excess risk from interaction*; RERI) (см. п. 12Б и вставку 8), также можно изучать с помощью моделей, основанных на относительных показателях.

Во многих случаях абсолютный риск (*absolute risk*), ассоциируемый с воздействием, более интересен, чем относительный риск. Например, если в фокусе исследования побочные эффекты лекарственного средства, интересным может быть число дополнительных случаев в единицу времени приема этого средства (дни, недели или годы). В примере выше указано дополнительное число случаев рака молочной железы на 1000 женщин, использовавших заместительную гормональную терапию в течение 10 лет [163]. Такие показатели, как дополнительный риск (*attributable risk*) или дополнительная доля риска в популяции (*population attributable fraction*), могут быть полезны для измерения числа случаев заболевания, которые можно было предотвратить, устранив воздействие. Их необходимо представлять вместе с показателями статистической неопределенности (например, с доверительными интервалами, как в примере выше). Авторы должны знать о значимых допущениях, сделанных в этом контексте, включая причинно-следственные связи между фактором риска и заболеванием (см. вставку 7) [171]. Вследствие семантической неоднозначности и сложности авторы должны подробно сообщить, какие методы использовались для расчета дополнительных рисков, приписываемых исследуемому фактору, в идеале — с указанием используемых формул [172].

Недавний обзор аннотаций 222 статей, опубликованных в ведущих медицинских журналах, показал, что в большинстве (62%) аннотаций рандомизированных исследований в качестве показателя отношения были указаны абсолютные риски. Что касается когортных исследований, такие данные были представлены в гораздо меньшем количестве статей (21%) [173]. Обзор информации в базе данных MEDLINE за период 1966–1997 гг. показал, что дополнительные риски (*attributable risks*) указывались в названии или аннотации статей 619 раз в сравнении с относительным риском или отношением шансов, упоминаемыми 18 955 раз, т.е. в соотношении 1 : 31 [174].

Вставка 7. Показатели связи, эффекта и влияния

Наблюдательные исследования могут проводиться исключительно для описания масштаба и распространения отклонений в состоянии здоровья среди населения. В таких исследованиях могут изучать количество людей, у которых есть определенное заболевание в конкретный момент времени (распространенность) или у которых заболевание развивается в течение определенного периода (инцидентность, заболеваемость). Заболеваемость может быть выражена долей лиц, у которых развивается заболевание (кумулятивная инцидентность; *cumulative incidence*), или показателем частоты событий в единицах человеко-времени (относительная инцидентность; *incidence rate*). Для описания различных типов инцидентности среди прочего используются такие термины, как смертность (*mortality rate*), рождаемость (*birth rate*), пораженность (*attack rate*) и летальность (*case fatality rate*). Аналогичным образом при описании различных типов распространенности говорят о распространенности заболевания в конкретный момент времени (*point prevalence*), за определенный период времени (*period prevalence*), в течение года (*annual prevalence*) или в течение жизни (*lifetime prevalence*) [30].

В других наблюдательных исследованиях рассматриваются причинно-следственные связи. Их целью является сравнение рисков, частоты или распространенности интересующего события между теми, кто подвергается или не подвергается воздействию исследуемого фактора риска. В этих исследованиях часто оценивается «относительный риск», который может означать отношения рисков (отношения показателей кумулятивной инцидентности; *ratios of cumulative incidences*), а также отношения частоты событий в единицах человеко-времени (отношения показателей относительной инцидентности; *ratios of incidence rates*). В исследования «случай-контроль» включают только часть исходной популяции («контроли»). Результаты выражаются как отношение шансов воздействия предполагаемой причины заболевания между «случаями» и «контролями». Отношение шансов (*odds ratio*) позволяет оценивать отношение рисков или частот в зависимости от выборки «случаев» и «контролей» (см. вставку 1) [175, 176]. В некоторых случаях может быть полезным определение отношения распространенности (*prevalence ratio*) или отношения шансов распространенности (*prevalence odds ratio*) исходя из данных одномоментных исследований [177].

Часто бывает полезным как относительное, так и абсолютное выражение результатов. Например, в исследовании, проведенном с участием британских врачей-мужчин, смертность от рака легких за 50 лет наблюдения составила 249 заболевших на 100 000 человек в год среди курильщиков по сравнению с 17 заболевшими на 100 000 человек в год сре-

ди некурящих, при этом отношение частот составляло 14,6 (249/17) [178]. Для случаев ишемической болезни сердца (ИБС) соответствующие показатели составили 1001 и 619 заболевших на 100 000 человек в год, отношение частот — 1,61 (1001/619). Выходит, что курение оказывает гораздо большее влияние на смертность от рака, чем от ИБС. Но если рассматривать абсолютный эффект курения, показатели будут другими. Разница в значениях относительной инцидентности составила 232 заболевших на 100 000 человек в год (249/17) для рака легких и 382 случая заболевания (1001/619) для ИБС. Таким образом, среди курящих докторов курение с большей вероятностью было причиной смерти от ИБС, чем от рака легких.

Какое бремя болезни в популяции можно предотвратить, устранив воздействие? Согласно глобальным оценкам, 91% всех случаев рака легких, 40% ИБС и 33% всех летальных исходов среди мужчин в 2000 г. были вызваны курением [179]. Дополнительная доля риска в популяции (*population attributable fraction*), как правило, оценивается как доля случаев, возникших под определенным воздействием, но существуют нескольких концепций определения этого показателя (без единой терминологии). Более того, иногда используются неверные подходы для коррекции величины дополнительного риска с учетом других факторов [172, 180]. Как следствие, относительные показатели указывают на значимость связей и наиболее полезны в этиологических исследованиях (*etiologic research*). Если причинно-следственная связь с воздействием задокументирована, а связи (*associations*) интерпретируются как эффекты, полученные оценки могут быть преобразованы из относительного риска в абсолютный, что позволит определить возможное влияние политики в области общественного здравоохранения (см. п. 16В) [181]. Однако в этом контексте авторам следует приводить достаточно убедительные доводы [171] и с большой осторожностью подходить к выбору концепции и методов анализа данных, подходящих к конкретной ситуации.

17. Другие анализы. Если проводились другие виды анализа, например, анализы в подгруппах, взаимодействий, чувствительности, укажите это

Пример 1

См. табл. 9.

Пример 2

См. табл. 10.

Пояснение

В дополнение к основному анализу в наблюдательных исследованиях часто используются и другие: анализ в специфических подгруппах, возможных взаимодействий между факторами риска, расчет дополнительных

Таблица 9. Анализ связи между приемом оральных контрацептивов, лейденской мутацией фактора V и риском возникновения венозной тромбоземболии ([182] с изменениями [183])

Table 9. Analysis of Oral Contraceptive Use, Presence of Factor V Leiden Allele, and Risk for Venous Thromboembolism (modified from [182] by [183])

Лейденская мутация фактора V	Оральные контрацептивы	Количество пациентов	Количество контролей	Отношение шансов
Да	Да	25	2	34,7
Да	Нет	10	4	6,9
Нет	Да	84	63	3,7
Нет	Нет	36	100	1 (референсная величина)

Таблица 10. Чувствительность отношения частоты сердечно-сосудистых исходов к неизмеренному вмешивающемуся фактору (адаптировано из [184])

Table 10. Sensitivity of the Rate Ratio for Cardiovascular Outcome to an Unmeasured Confounder (adapted from [184])

Распространенность неизмеренного бинарного вмешивающегося фактора в группе воздействия, %	Распространенность неизмеренного бинарного вмешивающегося фактора в группе сравнения, %	Отношение частоты для неизмеренного бинарного вмешивающегося фактора	Отношение частоты сильного воздействия (95% ДИ)*
90	10	1,5	1,20 (1,01–1,42)
90	50	1,5	1,43 (1,22–1,67)
50	10	1,5	1,39 (1,18–1,63)
90	10	2	0,96 (0,81–1,13)
90	50	2	1,27 (1,11–1,45)
50	10	2	1,21 (1,03–1,42)
90	50	3	1,18 (1,01–1,38)
50	10	3	0,99 (0,85–1,16)
90	50	5	1,08 (0,85–1,26)

Примечание. <*> — скорректировано с учетом возраста, пола, приема препаратов, назначенных при болезнях сердечно-сосудистой системы, неизмеренного бинарного вмешивающегося фактора. ДИ — доверительный интервал.

Note. <*> — Adjusted for age, sex, cardiovascular drug use, and unmeasured binary confounder. CI (ДИ) — confidence interval.

рисков (*attributable risks*), анализ чувствительности (*sensitivity analyses*) к смещениям при использовании альтернативных определений изучаемых переменных.

Ведется обсуждение опасностей, связанных с анализом в подгруппах и множественным анализом в целом [4, 104]. По нашему мнению, слишком велико стремление искать доказательств в связях, обнаруживаемых при анализе в подгруппах, когда результаты исследования в целом свидетельствуют о малом эффекте или полном его отсутствии. С другой стороны, имеет смысл изучить, сохраняется ли обнаруженная при основном анализе связь в заранее определенных подгруппах, особенно когда исследование достаточно велико, и данных по подгруппам достаточно много. Второй часто обсуждаемый вопрос — возникновение интереса к анализу в подгруппах в процессе анализа данных. Такие подгруппы могут стать источником важных находок, но иногда возникают случайно. Некоторые утверждают, что информировать читателя обо всех проведенных анализах в подгруппах не представляется ни возможным, ни целесообразным, поскольку только будущий анализ других данных покажет, в какой степени ранние захватывающие результаты выдержат испытание временем [9]. Мы советуем авторам указывать, какие анализы были запланированы, а какие нет (см. п. 4, 12Б и 20). Это позволит читателям оценить последствия множественного анализа и принять в расчет место исследования на всем протяжении от открытия до подтверждения или опровержения.

Третий предмет дискуссий — как оценивать совместный эффект и взаимодействие между факторами риска: с помощью аддитивной или мультипликативной шкалы? Или шкала должна определяться подходящей наилучшим образом статистической моделью (см. п. 12Б и вставку 8). Целесообразно описать и отдельные эффекты каждого воздействия, и совместный эффект, если возможно, в таблице, как в первом примере выше [183] или в исследовании I. Martinelli и соавт. [185]. Такая таблица дает читателю достаточно информации для оценки аддитивного и мультипликативного взаимодействия (см. вставку 8). Доверительные интервалы для отдельных и совместных эффектов помогут читателям судить о силе данных (*strength of the data*). Кроме того, доверительные

интервалы для показателей взаимодействия, таких как избыточный относительный риск (*relative excess risk from interaction*; RERI), относятся к тестам взаимодействия или однородности. Периодически возникающая проблема — использование авторами *p*-значений при сравнении подгрупп, что приводит к ошибочным заявлениям о модификаторе эффекта (*effect modifier*). Например, связь, статистически значимая в одной категории (скажем, у мужчин), но незначимая в другой (у женщин), сама по себе не доказывает модификацию эффекта. Точно так же доверительные интервалы для каждой точечной оценки иногда неправильно используются для вывода о том, что взаимодействие отсутствует, когда интервалы частично перекрываются. Более достоверные выводы можно сделать, если напрямую изучить, различается ли величина связи между подгруппами.

Анализ чувствительности полезен при изучении влияния выбора, сделанного в статистических анализах, или при выяснении устойчивости полученных результатов к отсутствующим данным и возможным систематическим ошибкам (см. п. 12Б). Авторам необходимо определиться, насколько подробно стоит освещать такие анализы в своих работах. Если предпринималось много анализов чувствительности, детальное их описание может не иметь практического смысла. Иногда достаточно сообщить, что такой анализ был проведен и что он согласуется с основными представленными результатами. Большая детальность уместна, если изучается очень важный вопрос или если оценки эффекта значительно варьируют [59, 186].

S.J. Росок и соавт. обнаружили, что 43 статьи из 73 с отчетами о наблюдательных исследованиях включали результаты анализа, проведенного в подгруппах. Большинство авторов сообщали о различиях между группами, и лишь в 8 статьях была представлена формальная оценка взаимодействия (см. п. 12Б) [4].

Вставка 8. Взаимодействие (модификация эффекта): анализ совместных эффектов

Взаимодействие возникает, если связь (*association*) воздействия с риском заболевания меняется при наличии другого воздействия (*exposure*). Одна из проблем оценки и описания взаимодействий заключается в том,

что эффект воздействия можно измерить двумя способами: как относительный риск (или отношение частот в единицах человека-времени) или как разность рисков (или разность частот). Использование относительного риска приводит к мультипликативной модели, в то время как использование разницы рисков соответствует аддитивной модели [187, 188]. Иногда проводится различие между «статистическим взаимодействием», которое может быть отклонением от мультипликативной или аддитивной модели, и «биологическим взаимодействием», которое измеряется отклонением от аддитивной модели [189]. Однако ни аддитивные, ни мультипликативные модели не указывают на конкретный биологический механизм взаимодействия.

Независимо от выбора модели важно понять, в какой степени совместный эффект двух факторов воздействия отличается от эффектов каждого из них (при отсутствии другого воздействия). HuGENet (Human Genomic Epidemiology Network) предложила схему прозрачного представления отдельных и совместных эффектов, позволяющую оценить различные типы взаимодействия [183]. Данные исследования о пероральных контрацептивах и лейденской мутации фактора V [182] были использованы для объяснения этого предложения, и этот пример также используется в п. 17. Каждый из двух представленных факторов в отдельности повышает риск возникновения венозного тромбоза. Вычисление их отдельных и совместных эффектов приводится в таблице 2×4 (см. пример 1 в п. 17), где отношение шансов со значением 1 означает исходный уровень для женщин без лейденской мутации, не принимающих оральные контрацептивы.

Сложность состоит в том, что некоторые дизайны, такие как исследование «случай-контроль», и некоторые статистические модели (логистическая регрессия или модель пропорциональных рисков Кокса) оценивают относительные риски (или отношение частот) и, по сути, ведут к мультипликативному моделированию. В этом случае относительные риски можно вычислить по аддитивной шкале. В примере 1 к п. 17 отношения шансов для каждого фактора равны 3,7 и 6,9, при одновременном действии двух факторов шансы равны 34,7. Если анализировать эти данные в рамках мультипликативной модели, ожидается, что общее отношение шансов составит 25,7 ($3,7 \times 6,9$). Наблюдаемый совместный эффект (34,7) в 1,4 раза превышает ожидаемый результат, определяемый по мультипликативной шкале ($34,7/25,7$). Эта величина (1,4) представляет собой отношение шансов мультипликативного взаимодействия (*odds ratio of the multiplicative interaction*). Такое же значение было бы получено при вычислении антилогарифма коэффициента взаимодействия в модели логистической регрессии. Согласно аддитивной шкале, отношение шансов при одновременном действии двух факторов составит 9,6 ($3,7 + 6,9 - 1$). Наблюдаемый совместный эффект двух факторов значительно отличается от аддитивного эффекта: разница составляет 25,1 ($34,7 - 9,6$). Если отношения шансов интерпретируются как относительные риски (или отношения частот), последняя величина (25,1) представляет собой значение относительного избыточного риска вследствие взаимодействия (*relative excess risk from interaction*; RERI) [190]. Это легко понять, если представить, что референсное значение (отношение шансов, равное 1) представляет собой исходную частоту развития венозного тромбоза, т.е., скажем, 1 случай на 10 000 человеко-лет, которая затем увеличивается при наличии отдельных и совместных воздействий.

Обсуждение

Раздел «Обсуждение» посвящен ключевым вопросам достоверности (*validity*) и смысла результатов исследования [191]. Исследования показали, что в этом разделе часто преобладают неполные и предвзятые оценки результатов исследования и их значения, а также риторика, поддерживающая находки авторов [192, 193]. Структурирование раздела «Обсуждение» поможет авторам избежать необоснованных предположений и переоценки результатов и вместе с тем направит читателей по тексту [194, 195]. К примеру, в журнале *Annals of Internal Medicine* [196] приводятся рекомендации по структурированию раздела «Обсуждение» согласно следующей схеме: 1) краткое изложение основных результатов, 2) рассмотрение возможных механизмов и объяснений, 3) сравнение с релевантными результатами других опубликованных исследований, 4) ограничения исследования, 5) краткое обобщение значения работы для практики и науки. Другие авторы внесли похожие предложения [191, 194]. Разделы статьи с рекомендациями для дальнейших исследований и обсуждением ограничений исследования должны быть тесно связаны друг с другом. Исследователи должны предложить способы, используя которые в последующих исследованиях можно достичь лучших результатов, а не вежливо заметить: «Необходимы дальнейшие исследования» [197, 198]. Мы предлагаем структурировать раздел «Обсуждение», используя подходящие подзаголовки.

18. Основные результаты. Обобщите основные результаты, ссылаясь на цели исследования

Пример

«Мы выдвинули гипотезу, согласно которой принадлежность к этническому меньшинству может быть связана с относительно высоким уровнем факторов риска сердечно-сосудистых заболеваний, и эта связь в значительной мере может быть приписана социально-экономическому статусу (СЭС). Наша гипотеза не подтвердилась. После поправки на возраст и СЭС сохранялась высокая статистически значимая разница по индексу массы тела, артериальному давлению, наличию сахарного диабета и низкой физической активности между белыми женщинами, афроамериканками и американками мексиканского происхождения. Кроме того, мы обнаружили значительную разницу уровней факторов риска сердечно-сосудистых заболеваний в зависимости от СЭС. Эти показатели иллюстрируют высокий риск как для женщин, относящихся к этническим меньшинствам, так и для белых женщин с низким СЭС» [199].

Пояснение

Хорошая практика — начинать обсуждение с резюмирования основных результатов исследования. Резюме напоминает читателям основные находки и помогает им оценить, подкреплены ли дальнейшие интерпретации и разъяснения авторов результатами работы.

19. Ограничения. Обсудите ограничения исследования с анализом источников потенциальных систематических ошибок или неточностей. Обсудите направление действия и значимость потенциальных систематических ошибок

Пример

«Поскольку распространенность консультирования возрастает с увеличением степени ожирения, мы можем переоценивать истинную распространенность.

Телефонный опрос также может приводить к завышению истинной распространенности консультирования. Хотя среди лиц, не имеющих доступа к телефонной связи, уровень ожирения схож с таковым, наблюдающимся среди тех, у кого такой доступ есть, отсутствие телефона имеет тенденцию быть связанным с более низким уровнем образования. Связь этого фактора с более низким уровнем консультирования продемонстрирована в нашем исследовании. Кроме того, возможно возникновение систематической ошибки из-за отказа некоторых лиц участвовать в опросе или отвечать на вопросы о своей массе тела. Более того, поскольку данные собирали одномоментно (*cross-sectionally*), мы не можем сделать вывод, что консультирование предшествовало попыткам пациентов снизить массу тела» [200].

Пояснение

Выявление и обсуждение ограничений исследования — необходимая часть научной публикации. Важно не только выявить источники систематических ошибок и искажений, вызванных действием вмешивающихся факторов (*confounding*), способных повлиять на результаты, но и обсудить относительную важность различных систематических ошибок, в том числе наиболее вероятное направление и величину любых возможных смещений (см. п. 9 и вставку 3).

Авторам следует также обсудить любые неточности результатов. Неточность возрастает в связи с различными аспектами исследования: размером исследования (*study size*) (п. 10), измерением воздействия, вмешивающихся факторов и исходов (п. 8). Неспособность точно измерить истинную величину воздействия ведет к возникновению смещения в направлении отсутствия эффекта (*bias towards unity*): чем менее точно измерены факторы риска, тем больше систематическая ошибка. Этот эффект описан как «аттенуация» (“*attenuation*”) [201, 202] или недавно как «систематическая ошибка регрессионного разведения» (“*regression dilution bias*”) [203]. Однако если коррелирующие факторы риска измеряются с разной степенью неточности, связанный с ними скорректированный относительный риск может быть смещен как в направлении единицы, так и от нее [204–206].

При обсуждении ограничений авторы могут сравнить свою работу с данными литературы с точки зрения достоверности (*validity*), обобщаемости (*generalisability*) и точности (*precision*). При таком подходе каждое исследование может рассматриваться как вклад в литературу, а не только как самостоятельная основа для выводов и действий [207]. Удивительно, но обсуждение важных ограничений исследования иногда не включается в опубликованные отчеты. Опрос авторов, опубликовавших результаты оригинальных исследований в журнале *The Lancet*, показал, что о важных недостатках исследования они сообщали в опросниках, но не в опубликованных статьях [192].

20. Интерпретация результатов.

Дайте осторожную, но достаточно полную интерпретацию результатов с учетом целей, ограничений, многочисленных анализов, результатов схожих исследований и других уместных свидетельств

Пример

«Любые объяснения связи между смертью от инфаркта миокарда и приемом оральных контрацептивов второго поколения должны носить характер гипотез. Не опубликованы данные, свидетельствующие о прямом

биологическом механизме такой связи, и не было других эпидемиологических исследований с подобными результатами. <...> Повышение абсолютного риска очень невелико и, возможно, объясняется главным образом курением. Из-за отсутствия прямых доказательств и из-за того, что анализ основан на относительно небольшом количестве данных, необходимо больше доказательств по этому вопросу. Мы не рекомендуем изменять существующую ныне практику назначения лекарственных препаратов на основании наших результатов» [120].

Пояснение

Суть раздела «Обсуждение» — интерпретация результатов исследования. Слишком смелая интерпретация — частое явление, свойственное человеческой природе: даже когда мы прикладываем большие усилия, стараясь дать объективную оценку, рецензенты справедливо указывают нам, что в некоторых аспектах мы зашли слишком далеко. При объяснении полученных результатов авторы должны принимать во внимание характер исследования на протяжении от открытия до его проверки, возможные источники систематических ошибок, в том числе выбывание участников из-под наблюдения и отказ от участия (см. п. 9, 12 и 19). Должное внимание должно быть уделено влиянию вмешивающихся факторов (п. 16А), результатам анализов чувствительности, вопросам множественного анализа и анализов в подгруппах (п. 17). Авторы должны также уделить внимание остаточному искажению (*residual confounding*) с учетом неизмеренных переменных или неточностей измерения вмешивающихся факторов. Например, социально-экономический статус связан со многими важными для здоровья исходами и часто различается между сравниваемыми группами. Переменные, используемые для измерения этого статуса (доход, образование, профессия) — всего лишь суррогаты других, неопределенных и неизмеренных, воздействий, и истинный вмешивающийся фактор по определению будет измерен с ошибкой [208]. Авторы должны обратить внимание на действительный диапазон неопределенности оценки (*real range of uncertainty*), который шире, чем статистическая неопределенность (*statistical uncertainty*), отражаемая в доверительном интервале. Последний не принимает в расчет неточности, связанные с дизайном исследования, его проведением и методами измерения [209].

Bradford Hill предложил в 1965 г. ряд критериев, которые могут быть полезными для размышлений и выводов о причинных связях [164]. Насколько сильна связь с воздействием? Предшествует ли воздействию развитию заболевания? Наблюдается ли выявленная связь в других исследованиях и условиях? Поддерживаются ли полученные результаты данными экспериментальных исследований, в том числе лабораторных, и исследований на животных? Насколько специфичен вызываемый воздействием эффект и является ли он дозозависимым? Является ли связь биологически правдоподобной? Однако эти критерии не должны применяться механически. Некоторые, например, считают, что относительный риск ниже 2 или 3 необходимо игнорировать [210, 211]. Это суждение противоречит точке зрения, высказанной J. Cornfield и соавт., о значении больших относительных рисков (см. п. 12Б) [127]. Хотя причинная связь более вероятна, если относительный риск составляет 9, из этого не следует, что все, что ниже 3, неправдоподобно. Например, небольшое повышение риска лейкемии у детей после внутриутробного облучения заслуживает доверия, поскольку объясняет возникновение побочного эффекта медицинской процедуры, для которого нет иных

очевидных причин [212]. Более того, канцерогенное влияние радиации считается доказанным. Удвоение риска рака яичников при съедании от 2 до 4 яиц в неделю не должно сразу же считаться убедительным, поскольку диета связана с большим числом факторов риска, обусловленных образом жизни и социально-экономическим статусом [213]. В противоположность этому правдоподобие широко обсуждаемых эпидемиологических данных о разнице в риске тромбоза между различными типами оральных контрацептивов стало еще более убедительным при обнаружении различий в коагуляции в рандомизированном перекрестном испытании (*randomised crossover trial*) [214]. Всегда следует включать обсуждение уже существующих доказательств, полученных в ходе других исследований, что особенно важно при обнаружении небольшого повышения риска. Авторы должны поместить свою работу в контекст схожих исследований и объяснить, как новые данные влияют на существующий комплекс доказательств, в идеале обращаясь к результатам систематического обзора.

21. Обобщаемость. Обсудите обобщаемость (внешнюю достоверность) результатов исследования

Пример

«Насколько применимы наши оценки к другим ВИЧ-1-инфицированным пациентам? Это важный вопрос, поскольку точность прогностической модели снижается, когда в нее включаются иные данные, отличающиеся от использованных при разработке. Для решения этой проблемы мы ограничили сложность модели и выбрали такую, которая обладает наилучшей обобщаемостью с учетом когорты, исключенных из процедуры оценки. Наша база данных включала информацию о пациентах из множества стран Европы и Северной Америки, получающих лечение в различных условиях. Спектр пациентов широк: мужчины и женщины, от подростков до пожилых людей. Основные категории воздействия представлены в достаточной мере. Тяжесть иммунодефицита в начале наблюдения варьирует от не поддающейся измерению до очень выраженной, вирусная нагрузка — от неопределимой до очень высокой» [215].

Пояснение

Обобщаемость (*generalisability*), также называемая внешней достоверностью (*external validity*) — мера применимости результатов исследования в других условиях [216]. Она не существует сама по себе, этот термин имеет смысл только в отношении четко определенных состояний [217]. Могут ли результаты быть применены к индивидуумам, группам или популяциям, отличающимся от участников исследования по полу, возрасту, этнической принадлежности, тяжести заболевания и наличию сопутствующих болезней? Сравнимы ли природа и уровень воздействия, а также определения исходов с теми, что наблюдаются в других условиях или популяциях? Если сбор данных в продолжительных когортных исследованиях (*longitudinal studies*) происходил на протяжении многих лет, остаются ли эти данные актуальными сегодня? Применимы ли результаты исследований системы здравоохранения в одной стране к системам здравоохранения других стран?

Вопрос внешней достоверности результатов исследования часто может быть решен с учетом сведений об условиях исследования, характеристиках участников исследования, воздействиях и исходах. Таким образом, крайне важно, чтобы авторы предоставляли читателям

адекватную информацию об условиях и месте проведения исследования, критериях отбора, воздействиях и способах их измерения, об определениях исходов, а также о периоде набора и последующего наблюдения. Степень неучастия и доля не подвергшихся воздействию участников, у которых развились исходы, также важны. Знание абсолютного риска и распространенности воздействия, которая часто варьирует от популяции к популяции, полезно при оценке применимости результатов в других условиях и популяциях (см. вставку 7).

Дополнительная информация

22. Финансирование. Укажите источники финансирования и роль спонсоров в настоящем исследовании и, если применимо, в оригинальном исследовании, на котором основывается представляемая статья

Пояснение

Некоторые журналы требуют от авторов раскрытия информации о наличии или отсутствии финансовых и иных конфликтов интересов [100, 218]. В нескольких работах продемонстрирована четкая связь между источником финансирования и выводами научных статей [219–222]. Если рандомизированное испытание финансируется коммерческой организацией, в выводах таких исследований значительно чаще (отношение шансов 5,3) рекомендуется использовать тестируемое лекарственное средство как препарат выбора [223]. Другие работы подтвердили влияние табачной и телекоммуникационной индустрии на финансируемые ими исследования [224–227]. Имеются также примеры неуместного влияния на выводы исследований спонсоров, в роли которых выступали правительственная или некоммерческая организации.

Авторы или спонсоры могут иметь конфликт интересов, который может влиять на дизайн исследования [228], выбор воздействий [228, 229], исходов [230], статистических методов [231], приводить к избирательной публикации информации об исходах [230] и исследовании в целом [232]. Следовательно, роль спонсоров должна быть детально описана: к какой части исследования они имели непосредственное отношение (например, выбор дизайна, сбор данных, анализ результатов, написание рукописи, принятие решения о публикации) [100]. Другие источники неуместного влияния включают работодателей (например, администрация университета для аффилированных научных работников и правительственные контролирующие органы, в частности политические деятели, для исследователей из государственных научных организаций), надзирающие органы, противников в судебной тяжбе, особые группы заинтересованных лиц.

ЗАКЛЮЧЕНИЕ

Цель STROBE — дать полезные рекомендации по составлению отчетов о наблюдательных исследованиях. Надлежащая отчетность выявляет сильные и слабые стороны исследования и способствует правильной интерпретации и применению результатов исследования. Рекомендации STROBE также могут быть полезны при планировании наблюдательных исследований и служить руководством для рецензентов и редакторов при работе над рукописью.

Наша пояснительная статья подчеркивает важность прозрачного и полного представления результатов исследований, включает в себя обоснование различных пунктов, включенных в контрольный перечень, а также примеры из опубликованных статей, иллюстрирующие нашу

точку зрения о том, что мы считаем качественными отчетами. Мы надеемся, что представленный здесь материал поможет авторам и редакторам в дальнейшем применении рекомендаций STROBE.

Мы подчеркиваем, что STROBE и другие рекомендации по описанию исследований [13, 233, 234] — документы, требующие постоянной оценки, уточнений и, если необходимо, изменений [235, 236]. Например, руководство CONSORT для описания рандомизированных исследований, проводимых в параллельных группах, было впервые представлено в середине 1990-х гг. [237]. С тех пор члены группы регулярно собираются для рассмотрения необходимости пересмотра рекомендаций; в 2001 г. появилась их обновленная версия [233], в настоящее время еще одна версия находится на стадии разработки. Аналогичным образом принципы, представленные в этой статье и в контрольном перечне STROBE, могут меняться по мере накопления новых доказательств и критических замечаний. Веб-сайт STROBE (<http://www.strobe-statement.org/>) предоставляет форум для обсуждения предложений по совершенствованию рекомендаций, настоящего пояснительного документа и другой дополнительной информации, способствующей повышению качества отчетов об эпидемиологических исследованиях.

Некоторые журналы в своих инструкциях для авторов просят следовать рекомендациям STROBE (актуальный список см. на <http://www.strobe-statement.org/>). Мы предлагаем другим журналам включить в инструкции для авторов руководство STROBE и сообщить об этом, связавшись с нами через наш веб-сайт. Журналы, публикующие рекомендации STROBE, предоставляют свободный доступ для читателей и всего биомедицинского сообщества.

ВЫРАЖЕНИЕ БЛАГОДАРНОСТИ

Мы благодарны Gerd Antes, Kay Dickersin, Shah Ebrahim, Richard Lilford за поддержку инициативы STROBE. Мы благодарны следующим институтам, на базе которых мы провели рабочие встречи: Institute of Social and Preventive Medicine, University of Bern (Швейцария); Department of Social Medicine, University of Bristol (Великобритания); London School of Hygiene & Tropical Medicine (Лондон, Великобритания); Nordic Cochrane Centre (Копенгаген, Дания); Centre for Statistics in Medicine (Оксфорд, Великобритания). Мы также благодарим четырех анонимных рецензентов, которые сопроводили рабочую версию этой статьи полезными комментариями.

ACKNOWLEDGMENTS

We are grateful to Gerd Antes, Kay Dickersin, Shah Ebrahim and Richard Lilford for supporting the STROBE Initiative. We are grateful to the following institutions that have hosted working meetings: Institute of Social and Preventive Medicine (ISPM), University of Bern, Switzerland; Department of Social Medicine, University of Bristol, UK; London School of Hygiene & Tropical Medicine, London, UK; Nordic Cochrane Centre, Copenhagen, Denmark; and Centre for Statistics in Medicine, Oxford, UK. We are grateful to four anonymous reviewers who provided helpful comments on a previous draft of this paper.

УЧАСТНИКИ ИНИЦИАТИВЫ STROBE

Следующие лица внесли вклад в создание и совершенствование руководства STROBE: Douglas G. Altman, Maria Blettner, Paolo Boffetta, Hermann Brenner, Genevieve Chene, Cyrus Cooper, George Davey-Smith, Erik von Elm, Matthias Egger, France Gagnon, Peter C. Gøtzsche, Philip

Greenland, Sander Greenland, Claire Infante-Rivard, John Ioannidis, Astrid James, Giselle Jones, Bruno Ledergerber, Julian Little, Margaret May, David Moher, Hooman Momen, Alfredo Morabia, Hal Morgenstern, Cynthia D. Mulrow, Fred Paccard, Stuart J. Pocock, Charles Poole, Martin Roosli, Dietrich Rothenbacher, Kenneth Rothman, Caroline Sabin, Willi Sauerbrei, Lale Say, James J. Schlesselman, Jonathan Sterne, Holly Syddall, Jan P. Vandenbroucke, Ian White, Susan Wieland, Hywel Williams, Guang Yong Zou.

CONTRIBUTORS TO THE STROBE INITIATIVE

The following persons have contributed to the content and elaboration of the STROBE Statement: Douglas G. Altman, Maria Blettner, Paolo Boffetta, Hermann Brenner, Genevieve Chène, Cyrus Cooper, George Davey-Smith, Erik von Elm, Matthias Egger, France Gagnon, Peter C. Gøtzsche, Philip Greenland, Sander Greenland, Claire Infante-Rivard, John Ioannidis, Astrid James, Giselle Jones, Bruno Ledergerber, Julian Little, Margaret May, David Moher, Hooman Momen, Alfredo Morabia, Hal Morgenstern, Cynthia D. Mulrow, Fred Paccard, Stuart J. Pocock, Charles Poole, Martin Rössli, Dietrich Rothenbacher, Kenneth Rothman, Caroline Sabin, Willi Sauerbrei, Lale Say, James J. Schlesselman, Jonathan Sterne, Holly Syddall, Jan P. Vandenbroucke, Ian White, Susan Wieland, Hywel Williams, Guang Yong Zou.

ВКЛАД АВТОРОВ

Все авторы внесли вклад в написание статьи. Jan P. Vandenbroucke, Erik von Elm, Douglas G. Altman, Peter C. Gøtzsche, Stuart J. Pocock, Matthias Egger — подготовка разделов первой версии рукописи. Erik von Elm — координация инициативы STROBE. Matthias Egger — инициировал STROBE и совместно с Erik von Elm организовал первый семинар для обсуждения STROBE.

AUTHOR CONTRIBUTIONS

All authors contributed to the writing of the paper. JPV, EvE, DGA, PCG, SJP, and ME wrote the first draft of different sections of the paper. EvE takes care of most of the practical coordination of STROBE. ME initiated STROBE and, together with EvE, organised the first workshop.

ИСТОЧНИК ФИНАНСИРОВАНИЯ

Первую рабочую встречу в рамках инициативы STROBE финансировал Европейский научный фонд (ESF). Дополнительное финансирование получено от Medical Research Council Health Services Research Collaboration и National Health Services Research & Development Methodology Programme. Спонсоры не влияли на планирование работы, сбор и анализ данных, подготовку рукописи или решение о ее публикации.

FINANCING SOURCE

The initial STROBE workshop was funded by the European Science Foundation (ESF). Additional funding was received from the Medical Research Council Health Services Research Collaboration and the National Health Services Research & Development Methodology Programme. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

РАСКРЫТИЕ ИНТЕРЕСОВ

Авторы статьи подтвердили отсутствие конфликта интересов, о котором необходимо сообщить.

DISCLOSURE OF INTERESTS

Not declared.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research. *BMJ*. 2004;328(7430):39–41. doi: <https://doi.org/10.1136/bmj.328.7430.39>
2. Funai EF, Rosenbush EJ, Lee MJ, Del Priore G. Distribution of study designs in four major US journals of obstetrics and gynecology. *Gynecol Obstet Invest*. 2001;51(1):8–11. doi: <https://doi.org/10.1159/000052882>
3. Scales CD, Norris RD, Peterson BL, et al. Clinical research and statistical methods in the urology literature. *J Urol*. 2005;174(4 Pt 1):1374–1379. doi: <https://doi.org/10.1097/01.ju.0000173640.91654.b5>
4. Pocock SJ, Collier TJ, Dandreo KJ, et al. Issues in the reporting of epidemiological studies: a survey of recent practice. *BMJ*. 2004;329(7471):883. doi: <https://doi.org/10.1136/bmj.38250.571088.55>
5. Tooth L, Ware R, Bain C, et al. Quality of reporting of observational longitudinal research. *Am J Epidemiol*. 2005;161(3):280–288. doi: <https://doi.org/10.1093/aje/kwi042>
6. Von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *Epidemiology*. 2007;18(6):800–804. doi: <https://doi.org/10.1097/EDE.0b013e3181577654>
7. Mihailovic A, Bell CM, Urbach DR. Users' guide to the surgical literature. Case-control studies in surgical journals. *Can J Surg*. 2005;48(2):148–151.
8. Rushton L. Reporting of occupational and environmental research: use and misuse of statistical and epidemiological methods. *Occup Environ Med*. 2000;57(1):1–9. doi: <https://doi.org/10.1136/oem.57.1.1>
9. Rothman KJ. No adjustments are needed for multiple comparisons. *Epidemiology*. 1990;1(1):43–46. doi: <https://doi.org/10.1097/00001648-199001000-00010>
10. Moonesinghe R, Khoury MJ, Janssens AC. Most published research findings are false-but a little replication goes a long way. *PLoS Med*. 2007;4(2):e28. doi: <https://doi.org/10.1371/journal.pmed.0040028>
11. Jenicek M. *Clinical Case Reporting. Evidence-Based Medicine*. Oxford: Butterworth-Heinemann; 1999. 117 p.
12. Vandenbroucke JP. In defense of case reports and case series. *Ann Intern Med*. 2001;134(4): 330–334. doi: <https://doi.org/10.7326/0003-4819-134-4-200102200-00017>
13. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med*. 2003;138(1):40–44. doi: <https://doi.org/10.7326/0003-4819-138-1-200301070-00010>
14. McShane LM, Altman DG, Sauerbrei W, et al. REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer*. 2005;93(4):387–391. doi: <https://doi.org/10.1038/sj.bjc.6602678>
15. Ioannidis JP, Gwinn M, Little J, et al. A road map for efficient and reliable human genome epidemiology. *Nat Genet*. 2006;38(1):3–5. doi: <https://doi.org/10.1038/ng0106-3>
16. Rodrigues L, Kirkwood BR. Case-control designs in the study of common diseases: updates on the demise of the rare disease assumption and the choice of sampling scheme for controls. *Int J Epidemiol*. 1990;19(1):205–213. doi: <https://doi.org/10.1093/ije/19.1.205>
17. Rothman KJ, Greenland S. Case-Control Studies. In: *Modern epidemiology*. Rothman KJ, Greenland S, eds. 2nd ed. Philadelphia: Lippincott Raven; 1998. pp. 93–114.
18. Forand SP. Leukaemia incidence among workers in the shoe and boot manufacturing industry: a case-control study. *Environ Health*. 2004;3(1):7. doi: <https://doi.org/10.1186/1476-069X-3-7>
19. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med*. 2000;342(25):1878–1886. doi: <https://doi.org/10.1056/NEJM200006223422506>
20. Gøtzsche PC, Harden A. *Searching for non-randomised studies*. Draft chapter 3. Cochrane Non-Randomised Studies Methods Group; 2002. Available online: <http://www.cochrane.dk/nrsmg>. Accessed on September 10, 2007.
21. Lohse N, Hansen AB, Pedersen G, et al. Survival of persons with and without HIV infection in Denmark, 1995–2005. *Ann Intern Med*. 2007;146(2):87–95. doi: <https://doi.org/10.7326/0003-4819-146-2-200701160-00003>
22. Information for authors. In: *American Journal of Epidemiology*. Available online: http://www.oxfordjournals.org/aje/for_authors/index.html. Accessed on September 10, 2007.
23. Haynes RB, Mulrow CD, Huth EJ, et al. More informative abstracts revisited. *Ann Intern Med*. 1990;113(1):69–76. doi: <https://doi.org/10.7326/0003-4819-113-1-69>
24. Taddio A, Pain T, Fassos FF, et al. Quality of nonstructured and structured abstracts of original research articles in the British Medical Journal, the Canadian Medical Association Journal and the Journal of the American Medical Association. *CMAJ*. 1994;150(10):1611–1615.
25. Hartley J, Sydes M. Which layout do you prefer? An analysis of readers' preferences for different typographic layouts of structured abstracts. *J Inform Sci*. 1996;22(1):27–37. doi: <https://doi.org/10.1177/016555159602200103>
26. Viner RM, Cole TJ. Adult socioeconomic, educational, social, and psychological outcomes of childhood obesity: a national birth cohort study. *BMJ*. 2005;330(7504):1354. doi: <https://doi.org/10.1136/bmj.38453.422049.E0>
27. McCauley J, Kern DE, Kolodner K, et al. The “battering syndrome”: prevalence and clinical characteristics of domestic violence in primary care internal medicine practices. *Ann Intern Med*. 1995;123(10):737–746. doi: <https://doi.org/10.7326/0003-4819-123-10-199511150-00001>
28. McEvoy SP, Stevenson MR, McCartt AT, et al. Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study. *BMJ*. 2005;331(7514):428. doi: <https://doi.org/10.1136/bmj.38537.397512.55>
29. Vandenbroucke JP. Prospective or retrospective: what's in a name? *BMJ*. 1991;302(6771):249–250. doi: <https://doi.org/10.1136/bmj.302.6771.249>
30. Last JM. *A Dictionary of Epidemiology*. New York: Oxford University Press; 2000.
31. Miettinen OS. *Theoretical Epidemiology: principles of occurrence research in medicine*. New York: Wiley; 1985. pp. 64–66.
32. Rothman KJ, Greenland S. Types of Epidemiologic Studies. In: *Modern epidemiology*. Rothman KJ, Greenland S, eds. 2nd ed. Lippincott Raven; 1998. pp. 74–75.
33. MacMahon B, Trichopoulos D. *Epidemiology, principles and methods*. 2nd ed. Boston: Little, Brown; 1996. 81 p. doi: [https://doi.org/10.1016/S0033-3506\(97\)00047-4](https://doi.org/10.1016/S0033-3506(97)00047-4)
34. Lilienfeld AM. *Foundations of Epidemiology*. New York: Oxford University Press; 1976.
35. Ridker PM, Hennekens CH, Lindpaintner K, et al. Mutation in the gene coding for coagulation factor V and the risk of myocardial infarction, stroke, and venous thrombosis in apparently healthy men. *N Engl J Med*. 1995;332(14):912–917. doi: <https://doi.org/10.1056/NEJM199504063321403>
36. Goodman KJ, O'Rourke K, Day RS, et al. Dynamics of Helicobacter pylori infection in a US-Mexico cohort during the first two years of life. *Int J Epidemiol*. 2005;34(6):1348–1355. doi: <https://doi.org/10.1093/ije/dyi152>
37. Altman DG, De Stavola BL, Love S.B, Stepniwska KA. Review of survival analyses published in cancer journals. *Br J Cancer*. 1995;72(2):511–518. doi: <https://doi.org/10.1038/bjc.1995.364>
38. Cerhan JR, Wallace RB, Folsom AR, et al. Transfusion history and cancer risk in older women. *Ann Intern Med*. 1993;119(1): 8–15. doi: <https://doi.org/10.7326/0003-4819-119-1-199307010-00002>
39. Freeman LE, Dennis LK, Lynch CF, et al. Toenail arsenic content and cutaneous melanoma in Iowa. *Am J Epidemiol*. 2004;160(7):679–687. doi: <https://doi.org/10.1093/aje/kwh267>
40. Canto JG, Allison JJ, Kiefe CI, et al. Relation of race and sex to the use of reperfusion therapy in Medicare beneficiaries with acute myocardial infarction. *N Engl J Med*. 2000;342(15):1094–1100. doi: <https://doi.org/10.1056/NEJM200004133421505>
41. Metztkor-Cotter E, Kletter Y, Avidor B, et al. Long-term serological analysis and clinical follow-up of patients with cat scratch disease. *Clin Infect Dis*. 2003;37(9):1149–1154. doi: <https://doi.org/10.1086/378738>
42. Johnson ES. Bias on withdrawing lost subjects from the analysis at the time of loss, in cohort mortality studies, and in follow-up methods. *J Occup Med*. 1990;32(3):250–254. doi: <https://doi.org/10.1097/00043764-199003000-00013>

43. Berkson J. Limitations of the application of fourfold table analysis to hospital data. *Biom Bull.* 1946;2(3):47. doi: <https://doi.org/10.2307/3002000>
44. Feinstein AR, Walter SD, Horwitz RI. An analysis of Berkson's bias in case-control studies. *J Chronic Dis.* 1986;39(7):495–504. doi: [https://doi.org/10.1016/0021-9681\(86\)90194-3](https://doi.org/10.1016/0021-9681(86)90194-3)
45. Jick H, Vessey MP. Case-control studies in the evaluation of drug-induced illness. *Am J Epidemiol.* 1978;107(1):1–7. doi: <https://doi.org/10.1093/oxfordjournals.aje.a112502>
46. Hackam DG, Mamdani M, Li P, Redelmeier DA. Statins and sepsis in patients with cardiovascular disease: a population-based cohort analysis. *Lancet.* 2006;367(9508):413–418. doi: [https://doi.org/10.1016/S0140-6736\(06\)68041-0](https://doi.org/10.1016/S0140-6736(06)68041-0)
47. Smeeth L, Cook C, Fombonne E, et al. MMR vaccination and pervasive developmental disorders: a case-control study. *Lancet.* 2004;364(9438):963–969. doi: [https://doi.org/10.1016/S0140-6736\(04\)17020-7](https://doi.org/10.1016/S0140-6736(04)17020-7)
48. Costanza MC. Matching. *Prev Med.* 1995;24(5):425–433. doi: <https://doi.org/10.1006/pmed.1995.1069>
49. Sturmer T, Brenner H. Flexible matching strategies to increase power and efficiency to detect and estimate gene-environment interactions in case-control studies. *Am J Epidemiol.* 2002;155(7):593–602. doi: <https://doi.org/10.1093/aje/155.7.593>
50. Rothman KJ, Greenland S. Matching. In: *Modern epidemiology*. Rothman KJ, Greenland S, eds. 2nd ed. Lippincott Raven; 1998. p. 147–161.
51. Szklo MF, Nieto J. *Epidemiology, Beyond the Basics*. Sudbury (MA): Jones and Bartlett; 2000. p. 40–51.
52. Cole P, MacMahon B. Attributable risk percent in case-control studies. *Br J Prev Soc Med.* 1971;25(4):242–244. doi: <https://doi.org/10.1136/jech.25.4.242>
53. Gissler M, Hemminki E. The danger of overmatching in studies of the perinatal mortality and birthweight of infants born after assisted conception. *Eur J Obstet Gynecol Reprod Biol.* 1996;69(2):73–75. doi: [https://doi.org/10.1016/0301-2115\(95\)02517-0](https://doi.org/10.1016/0301-2115(95)02517-0)
54. Gefeller O, Pfahlerberg A, Brenner H, Windeler J. An empirical investigation on matching in published case-control studies. *Eur J Epidemiol.* 1998;14(4):321–325. doi: <https://doi.org/10.1023/A:1007497104800>
55. Artama M, Ritvanen A, Gissler M, et al. Congenital structural anomalies in offspring of women with epilepsy—a population-based cohort study in Finland. *Int J Epidemiol.* 2006;35(2):280–287. doi: <https://doi.org/10.1093/ije/dyi234>
56. Ebrahim S. Cohorts, infants and children. *Int J Epidemiol.* 2004;33(6):1165–1166. doi: <https://doi.org/10.1093/ije/dyh368>
57. Walker M, Whincup PH, Shaper AG. The British Regional Heart Study 1975–2004. *Int J Epidemiol.* 2004;33(6):1185–1192. doi: <https://doi.org/10.1093/ije/dyh295>
58. Wieland S, Dickersin K. Selective exposure reporting and Medline indexing limited the search sensitivity for observational studies of the adverse effects of oral contraceptives. *J Clin Epidemiol.* 2005;58(6):560–567. doi: <https://doi.org/10.1016/j.jclinepi.2004.11.018>
59. Anderson HR, Atkinson RW, Peacock JL, et al. Ambient particulate matter and health effects: publication bias in studies of short-term associations. *Epidemiology.* 2005;16(2):155–163. doi: <https://doi.org/10.1097/01.ede.0000152528.22746.0f>
60. Winkelmayr WC, Stampfer MJ, Willett WC, Curhan GC. Habitual caffeine intake and the risk of hypertension in women. *JAMA.* 2005;294(18):2330–2335. doi: <https://doi.org/10.1001/jama.294.18.2330>
61. Lukanova A, Soderberg S, Kaaks R, et al. Serum adiponectin is not associated with risk of colorectal cancer. *Cancer Epidemiol Biomarkers Prev.* 2006;15(2):401–402. doi: <https://doi.org/10.1158/1055-9965.EPI-05-0836>
62. Becher H. The concept of residual confounding in regression models and some applications. *Stat Med.* 1992;11(13):1747–1758. doi: <https://doi.org/10.1002/sim.4780111308>
63. Brenner H, Blettner M. Controlling for continuous confounders in epidemiologic research. *Epidemiology.* 1997;8(4):429–434. doi: <https://doi.org/10.1097/00001648-199707000-00014>
64. Phillips MR, Yang G, Zhang Y, et al. Risk factors for suicide in China: a national case-control psychological autopsy study. *Lancet.* 2002;360(9347):1728–1736. doi: [https://doi.org/10.1016/S0140-6736\(02\)11681-3](https://doi.org/10.1016/S0140-6736(02)11681-3)
65. Pasquale LR, Kang JH, Manson JE, et al. Prospective study of type 2 diabetes mellitus and risk of primary open-angle glaucoma in women. *Ophthalmology.* 2006. Vol. 113, N 7. P. 1081–1086. doi: <https://doi.org/10.1016/j.ophtha.2006.01.066>
66. Craig SL, Feinstein AR. Antecedent therapy versus detection bias as causes of neoplastic multimorbidity. *Am J Clin Oncol.* 1999;22(1):51–56. doi: <https://doi.org/10.1097/00000421-199902000-00013>
67. Rogler LH, Mroczek DK, Fellows M, Loftus ST. The neglect of response bias in mental health research. *J Nerv Ment Dis.* 2001;189(3):182–187. doi: <https://doi.org/10.1097/00005053-200103000-00007>
68. Murphy EA. *The logic of medicine*. Baltimore: Johns Hopkins University Press; 1976.
69. Sackett DL. Bias in analytic research. *J Chronic Dis.* 1979;32(1-2):51–63. doi: [https://doi.org/10.1016/0021-9681\(79\)90012-2](https://doi.org/10.1016/0021-9681(79)90012-2)
70. Johannes CB, Crawford SL, McKinlay JB. Interviewer effects in a cohort study. Results from the Massachusetts Women's Health Study. *Am J Epidemiol.* 1997;146(5):429–438. doi: <https://doi.org/10.1093/oxfordjournals.aje.a009296>
71. Bloemkamp KW, Rosendaal FR, Buller HR, et al. Risk of venous thrombosis with use of current low-dose oral contraceptives is not explained by diagnostic suspicion and referral bias. *Arch Intern Med.* 1999;159(1):65–70. doi: <https://doi.org/10.1001/archinte.159.1.65>
72. Feinstein AR. *Clinical epidemiology: the architecture of clinical research*. Philadelphia: W.B. Saunders; 1985.
73. Yadon ZE, Rodrigues LC, Davies CR, Quigley MA. Indoor and peridomestic transmission of American cutaneous leishmaniasis in northwestern Argentina: a retrospective case-control study. *Am J Trop Med Hyg.* 2003;68(5):519–526. doi: <https://doi.org/10.4269/ajtmh.2003.68.519>
74. Anoop S, Saravanan B, Joseph A, et al. Maternal depression and low maternal intelligence as risk factors for malnutrition in children: a community based case-control study from South India. *Arch Dis Child.* 2004;89(4):325–329. doi: <https://doi.org/10.1136/adc.2002.009738>
75. Carlin JB, Doyle LW. Sample size. *J Paediatr Child Health.* 2002;38(3):300–304. doi: <https://doi.org/10.1046/j.1440-1754.2002.00855.x>
76. Rigby AS, Vail A. Statistical methods in epidemiology. II: A commonsense approach to sample size estimation. *Disabil Rehabil.* 1998;20(11):405–410. doi: <https://doi.org/10.3109/09638289809166102>
77. Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. *Lancet.* 2005;365(9467):1348–1353. doi: [https://doi.org/10.1016/S0140-6736\(05\)61034-3](https://doi.org/10.1016/S0140-6736(05)61034-3)
78. Drescher K, Timm J, Jockel KH. The design of case-control studies: the effect of confounding on sample size requirements. *Stat Med.* 1990;9(7):765–776. doi: <https://doi.org/10.1002/sim.4780090706>
79. Devine OJ, Smith JM. Estimating sample size for epidemiologic studies: the impact of ignoring exposure measurement uncertainty. *Stat Med.* 1998;17(12):1375–1389. doi: [https://doi.org/10.1002/\(SICI\)1097-0258\(19980630\)17:12<1375::AID-SIM857>3.0.CO;2-D](https://doi.org/10.1002/(SICI)1097-0258(19980630)17:12<1375::AID-SIM857>3.0.CO;2-D)
80. Linn S, Levi L, Grunau PD., et al. Effect measure modification and confounding of severe head injury mortality by age and multiple organ injury severity. *Ann Epidemiol.* 2007;17(2):142–147. doi: <https://doi.org/10.1016/j.annepidem.2006.08.004>
81. Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst.* 1994;86(11):829–835. doi: <https://doi.org/10.1093/jnci/86.11.829>
82. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med.* 2006;25(1):127–141. doi: <https://doi.org/10.1002/sim.2331>
83. Greenland S. Avoiding power loss associated with categorization and ordinal scores in dose-response and trend analysis. *Epidemiology.* 1995;6(4):450–454. doi: <https://doi.org/10.1097/00001648-199507000-00025>
84. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol.* 1999;28(5):964–974. doi: <https://doi.org/10.1093/ije/28.5.964>
85. MacCallum RC, Zhang S, Preacher KJ, Rucker D.D. On the practice of dichotomization of quantitative variables. *Psychol Methods.* 2002;7(1):19–40. doi: <https://doi.org/10.1037/1082-989X.7.1.19>

86. Altman D.G. Categorizing continuous variables. In: *Encyclopedia of biostatistics*. Armitage P, Colton T, eds. 2nd ed. Chichester: John Wiley; 2005. pp. 708–711. doi: <https://doi.org/10.1002/0470011815.b2a10012>
87. Cohen J. The cost of dichotomization. *Applied Psychological Measurement*. 1983;7(3):249–253. doi: <https://doi.org/10.1177/014662168300700301>
88. Zhao LP, Kolonel LN. Efficiency loss from categorizing quantitative exposures into qualitative exposures in case-control studies. *Am J Epidemiol*. 1992;136(4):464–474. doi: <https://doi.org/10.1093/oxfordjournals.aje.a116520>
89. Cochran WG. The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*. 1968;24(2):295–313. doi: <https://doi.org/10.2307/2528036>
90. Clayton D, Hills M. Models for dose-response. Chapter 25. In: *Statistical Models in Epidemiology*. Oxford: Oxford University Press; 1993. P. 249–260.
91. Cox DR. Note on grouping. *J Am Stat Assoc*. 1957;52(280):543–547. doi: <https://doi.org/10.1080/01621459.1957.10501411>
92. Il'yasova D, Hertz-Picciotto I, Peters U, et al. Choice of exposure scores for categorical regression in meta-analysis: a case study of a common problem. *Cancer Causes Control*. 2005;16(4):383–388. doi: <https://doi.org/10.1007/s10552-004-5025-x>
93. Berglund A, Alfredsson L, Cassidy JD, et al. The association between exposure to a rear-end collision and future neck or shoulder pain: a cohort study. *J Clin Epidemiol*. 2000;53(11):1089–1094. doi: [https://doi.org/10.1016/S0895-4356\(00\)00225-0](https://doi.org/10.1016/S0895-4356(00)00225-0)
94. Slama R, Werwatz A. Controlling for continuous confounding factors: non- and semiparametric approaches. *Rev Epidemiol Sante Publique*. 2005;53(Spec No 2):2S65–2S80. doi: [https://doi.org/10.1016/S0398-7620\(05\)84769-8](https://doi.org/10.1016/S0398-7620(05)84769-8)
95. Greenland S. Introduction to regression modelling. Chapter 21. In: *Modern epidemiology*. Rothman KJ, Greenland S, eds. 2nd ed. Lippincott Raven; 1998. pp. 401–432.
96. Thompson WD. Statistical analysis of case-control studies. *Epidemiol Rev*. 1994;16(1):33–50. doi: <https://doi.org/10.1093/oxfordjournals.epirev.a036143>
97. Schlesselman JJ. Logistic regression for case-control studies. Chapter 8.2. *Case-control studies: Design, conduct, analysis*. New York, Oxford: Oxford University Press; 1982. pp. 235–241.
98. Clayton D, Hills M. Choice and interpretation of models. Chapter 27. *Statistical Models in Epidemiology*. Oxford: Oxford University Press; 1993. pp. 271–281.
99. Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. *Br Med J*. 1983;286(6376):1489–1493. doi: <https://doi.org/10.1136/bmj.286.6376.1489>
100. International Committee of Medical Journal Editors Uniform requirements for manuscripts submitted to biomedical journals. *N Engl J Med*. 1997;336:309–315. doi: <https://doi.org/10.1056/NEJM199701233360422>
101. Mullner M, Matthews H, Altman D.G. Reporting on statistical methods to adjust for confounding: a cross-sectional survey. *Ann Intern Med*. 2002;136(2):122–126. doi: <https://doi.org/10.7326/0003-4819-136-2-200201150-00009>
102. Olsen J, Basso O. Residual confounding. *Am J Epidemiol*. 1999;149(3):290. doi: <https://doi.org/10.1093/oxfordjournals.aje.a009805>
103. Hallan S, de Mutsert R, Carlsen S, et al. Obesity, smoking, and physical inactivity as risk factors for CKD: are men more vulnerable? *Am J Kidney Dis*. 2006;47(3):396–405. doi: <https://doi.org/10.1053/j.ajkd.2005.11.027>
104. Gotzsche PC. Believability of relative risks and odds ratios in abstracts: cross sectional study. *BMJ*. 2006;333(7561):231–234. doi: <https://doi.org/10.1136/bmj.38895.410451.79>
105. Szklo MF, Nieto J. Communicating Results of Epidemiologic Studies. Chapter 9. In: *Epidemiology, Beyond the Basics*. Sudbury (MA): Jones and Bartlett; 2000. pp. 408–430.
106. Chandola T, Brunner E, Marmot M. Chronic stress at work and the metabolic syndrome: prospective study. *BMJ*. 2006;332(7540):521–525. doi: <https://doi.org/10.1136/bmj.38693.435301.80>
107. Vach W, Blettner M. Biased estimation of the odds ratio in case-control studies due to the use of ad hoc methods of correcting for missing values for confounding variables. *Am J Epidemiol*. 1991;134(8):895–907. doi: <https://doi.org/10.1093/oxfordjournals.aje.a116164>
108. Little RJ, Rubin DB. A taxonomy of missing-data methods. Chapter 1.4. In: *Statistical Analysis with Missing Data*. New York: Wiley; 2002. pp. 19–23. doi: <https://doi.org/10.1002/9781119013563>
109. Ware JH. Interpreting incomplete data in studies of diet and weight loss. *N Engl J Med*. 2003;348(21):2136–2137. doi: <https://doi.org/10.1056/NEJMe030054>
110. Rubin DB. Inference and missing data. *Biometrika*. 1976;63(3):581–592. doi: <https://doi.org/10.1093/biomet/63.3.581>
111. Schafer JL. *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall; 1997. doi: <https://doi.org/10.1201/9781439821862>
112. Lipsitz SR, Ibrahim JG, Chen MH, Peterson H. Non-ignorable missing covariates in generalized linear models. *Stat Med*. 1997;18(17-18):2435–2448. doi: [https://doi.org/10.1002/\(SICI\)1097-0258\(19990915/30\)18:17/18<2435::AID-SIM267>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1097-0258(19990915/30)18:17/18<2435::AID-SIM267>3.0.CO;2-B)
113. Rotnitzky A, Robins J. Analysis of semi-parametric regression models with non-ignorable non-response. *Stat Med*. 1997;16(1-3):81–102. doi: [https://doi.org/10.1002/\(SICI\)1097-0258\(19970115\)16:1<81::AID-SIM473>3.0.CO;2-0](https://doi.org/10.1002/(SICI)1097-0258(19970115)16:1<81::AID-SIM473>3.0.CO;2-0)
114. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley; 1987. doi: <https://doi.org/10.1002/9780470316696>
115. Barnard J, Meng XL. Applications of multiple imputation in medical studies: from AIDS to NHANES. *Stat Methods Med Res*. 1999;8(1):17–36. doi: <https://doi.org/10.1177/096228029900800103>
116. Braitstein P, Brinkhof MW, Dabis F, et al. Mortality of HIV-1-infected patients in the first year of antiretroviral therapy: comparison between low-income and high-income countries. *Lancet*. 2006;367(9513):817–824. doi: [https://doi.org/10.1016/S0140-6736\(06\)68337-2](https://doi.org/10.1016/S0140-6736(06)68337-2)
117. Purandare N, Burns A, Daly KJ, et al. Cerebral emboli as a potential cause of Alzheimer's disease and vascular dementia: case-control study. *BMJ*. 2006. Vol. 332, N 7550. P. 1119–1124. doi: <https://doi.org/10.1136/bmj.38814.696493.AE>
118. Steyn K, Gaziano TA, Bradshaw D, et al. Hypertension in South African adults: results from the Demographic and Health Survey, 1998. *J Hypertens*. 2001;19(10):1717–1725. doi: <https://doi.org/10.1097/00004872-200110000-00004>
119. Lohr SL. Design Effects. Chapter 7.5. In: *Sampling: Design and Analysis*. Pacific Grove (CA): Duxbury Press; 1999.
120. Dunn NR, Arscott A, Thorogood M. The relationship between use of oral contraceptives and myocardial infarction in young women with fatal outcome, compared to those who survive: results from the MICA case-control study. *Contraception*. 2001;63(2):65–69. doi: [https://doi.org/10.1016/S0010-7824\(01\)00172-X](https://doi.org/10.1016/S0010-7824(01)00172-X)
121. Rothman KJ, Greenland S. Basic Methods for Sensitivity Analysis and External Adjustment. In: *Modern epidemiology*. Rothman KJ, Greenland S, eds. 2nd ed. Lippincott Raven; 1998. pp. 343–357.
122. Custer B, Longstreth WT, Phillips LE, et al. Hormonal exposures and the risk of intracranial meningioma in women: a population-based case-control study. *BMC Cancer*. 2006;6:152. doi: <https://doi.org/10.1186/1471-2407-6-152>
123. Wakefield MA, Chaloupka FJ, Kaufman NJ, et al. Effect of restrictions on smoking at home, at school, and in public places on teenage smoking: cross sectional study. *BMJ*. 2000;321(7257):333–337. doi: <https://doi.org/10.1136/bmj.321.7257.333>
124. Greenland S. The impact of prior distributions for uncontrolled confounding and response bias: a case study of the relation of wire codes and magnetic fields to childhood leukemia. *J Am Stat Assoc*. 2003;98(461):47–54. doi: <https://doi.org/10.1198/01621450338861905>
125. Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology*. 2003;14(4):451–458. doi: <https://doi.org/10.1097/01.EDE.0000071419.41011.cf>
126. Phillips CV. Quantifying and reporting uncertainty from systematic errors. *Epidemiology*. 2003;14(4):459–466. doi: <https://doi.org/10.1097/01.ede.0000072106.65262.ae>
127. Cornfield J, Haenszel W, Hammond EC, et al. Smoking and lung cancer: recent evidence and a discussion of some questions. *J Natl Cancer Inst*. 1959;22(1):173–203.

128. Langholz B. Factors that explain the power line configuration wiring code-childhood leukemia association: what would they look like? *Bioelectromagnetics*. 2001;Suppl 5:S19–S31. doi: [https://doi.org/10.1002/1521-186x\(2001\)22:5+::aid-bem1021>3.3.co;2-9](https://doi.org/10.1002/1521-186x(2001)22:5+::aid-bem1021>3.3.co;2-9)
129. Eisner MD, Smith AK, Blanc PD. Bartenders' respiratory health after establishment of smoke-free bars and taverns. *JAMA*. 1998;280(22):1909–1914. doi: <https://doi.org/10.1001/jama.280.22.1909>
130. Dunne MP, Martin NG, Bailey JM, et al. Participation bias in a sexuality survey: psychological and behavioural characteristics of responders and non-responders. *Int J Epidemiol*. 1997;26(4):844–854. doi: <https://doi.org/10.1093/ije/26.4.844>
131. Schuz J, Kaatsch P, Kaletsch U, et al. Association of childhood cancer with factors related to pregnancy and birth. *Int J Epidemiol*. 1999;28(4):631–639. doi: <https://doi.org/10.1093/ije/28.4.631>
132. Cnattingius S, Zack M, Ekblom A, et al. Prenatal and neonatal risk factors for childhood myeloid leukemia. *Cancer Epidemiol Biomarkers Prev*. 1995;4(5):441–445.
133. Schuz J. Non-response bias as a likely cause of the association between young maternal age at the time of delivery and the risk of cancer in the offspring. *Paediatr Perinat Epidemiol*. 2003;17(1):106–112. doi: <https://doi.org/10.1046/j.1365-3016.2003.00460.x>
134. Slattery ML, Edwards SL, Caan BJ, et al. Response rates among control subjects in case-control studies. *Ann Epidemiol*. 1995;5(3):245–249. doi: [https://doi.org/10.1016/1047-2797\(94\)00113-8](https://doi.org/10.1016/1047-2797(94)00113-8)
135. Schulz KF, Grimes DA. Case-control studies: research in reverse. *Lancet*. 2002;359(9304):431–434. doi: [https://doi.org/10.1016/S0140-6736\(02\)07605-5](https://doi.org/10.1016/S0140-6736(02)07605-5)
136. Olson SH, Voigt LF, Begg CB, Weiss NS. Reporting participation in case-control studies. *Epidemiology*. 2002;13(2):123–126. doi: <https://doi.org/10.1097/00001648-200203000-00004>
137. Morton LM, Cahill J, Hartge P. Reporting participation in epidemiologic studies: a survey of practice. *Am J Epidemiol*. 2006;163(3):197–203. doi: <https://doi.org/10.1093/aje/kwj036>
138. Olson SH. Reported participation in case-control studies: changes over time. *Am J Epidemiol*. 2001;154(6):574–581. doi: <https://doi.org/10.1093/aje/154.6.574>
139. Sandler DP. On revealing what we'd rather hide: the problem of describing study participation. *Epidemiology*. 2002;13(2):117. doi: <https://doi.org/10.1097/00001648-200203000-00001>
140. Hepworth SJ, Schoemaker MJ, Muir KR, et al. Mobile phone use and risk of glioma in adults: case-control study. *BMJ*. 2006;332(7546):883–887. doi: <https://doi.org/10.1136/bmj.38720.687975.55>
141. Hay AD, Wilson A, Fahey T, Peters TJ. The duration of acute cough in pre-school children presenting to primary care: a prospective cohort study. *Fam Pract*. 2003;20(6):696–705. doi: <https://doi.org/10.1093/fampra/cmg613>
142. Egger M, Juni P, Bartlett C. Value of flow diagrams in reports of randomized controlled trials. *JAMA*. 2001;285(15):1996–1999. doi: <https://doi.org/10.1001/jama.285.15.1996>
143. Osella AR, Misciagna G, Guerra VM, et al. Hepatitis C virus (HCV) infection and liver-related mortality: a population-based cohort study in southern Italy. The Association for the Study of Liver Disease in Puglia. *Int J Epidemiol*. 2000;29(5):922–927. doi: <https://doi.org/10.1093/ije/29.5.922>
144. Dales LG, Ury HK. An improper use of statistical significance testing in studying covariables. *Int J Epidemiol*. 1978;7(4):373–375. doi: <https://doi.org/10.1093/ije/7.4.373>
145. Maldonado G, Greenland S. Simulation study of confounder-selection strategies. *Am J Epidemiol*. 1993;138(11):923–936. doi: <https://doi.org/10.1093/oxfordjournals.aje.a116813>
146. Tanis BC, van den Bosch MA, Kemmeren JM, et al. Oral contraceptives and the risk of myocardial infarction. *N Engl J Med*. 2001;345(25):1787–1793. doi: <https://doi.org/10.1056/NEJMoa003216>
147. Rothman KJ, Greenland S. Precision and Validity in Epidemiologic Studies. In: *Modern epidemiology*. Rothman KJ, Greenland S, eds. 2nd ed. Lippincott Raven; 1998. pp. 120–125.
148. Clark TG, Altman DG, De Stavola BL. Quantification of the completeness of follow-up. *Lancet*. 2002;359(9314):1309–1310. doi: [https://doi.org/10.1016/S0140-6736\(02\)08272-7](https://doi.org/10.1016/S0140-6736(02)08272-7)
149. Qiu C, Fratiglioni L, Karp A, et al. Occupational exposure to electromagnetic fields and risk of Alzheimer's disease. *Epidemiology*. 2004;15(6):687–694. doi: <https://doi.org/10.1097/01.ede.0000142147.49297.9d>
150. Kengeya-Kayondo JF, Kamali A, Nunn AJ, et al. Incidence of HIV-1 infection in adults and socio-demographic characteristics of seroconverters in a rural population in Uganda: 1990–1994. *Int J Epidemiol*. 1996;25(5):1077–1082. doi: <https://doi.org/10.1093/ije/25.5.1077>
151. Mastrangelo G, Fedeli U, Fadda E, et al. Increased risk of hepatocellular carcinoma and liver cirrhosis in vinyl chloride workers: synergistic effect of occupational exposure with alcohol intake. *Environ Health Perspect*. 2004;112(11):1188–1192. doi: <https://doi.org/10.1289/ehp.6972>
152. Salo PM, Arbes SJ, Sever M, et al. Exposure to *Alternaria alternata* in US homes is associated with asthma symptoms. *J Allergy Clin Immunol*. 2006;118(4):892–898. doi: <https://doi.org/10.1016/j.jaci.2006.07.037>
153. Pocock SJ, Clayton TC, Altman DG. Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. *Lancet*. 2002;359(9318):1686–1689. doi: [https://doi.org/10.1016/S0140-6736\(02\)08594-X](https://doi.org/10.1016/S0140-6736(02)08594-X)
154. Sasieni P. A note on the presentation of matched case-control data. *Stat Med*. 1992;11(5):617–620. doi: <https://doi.org/10.1002/sim.4780110506>
155. Lee GM, Neutra RR, Hristova L, et al. A nested case-control study of residential and personal magnetic field measures and miscarriages. *Epidemiology*. 2002;13(1):21–31. doi: <https://doi.org/10.1097/00001648-200201000-00005>
156. Tiihonen J, Walhbeck K, Lonnqvist J, et al. Effectiveness of antipsychotic treatments in a nationwide cohort of patients in community care after first hospitalisation due to schizophrenia and schizoaffective disorder: observational follow-up study. *BMJ*. 2006;333(7561):224. doi: <https://doi.org/10.1136/bmj.38881.382755.2F>
157. Christenfeld NJ, Sloan RP, Carroll D, Greenland S. Risk factors, confounding, and the illusion of statistical control. *Psychosom Med*. 2004;66(6):868–875. doi: <https://doi.org/10.1097/01.psy.0000140008.70959.41>
158. Smith GD, Phillips A. Declaring independence: why we should be cautious. *J Epidemiol Community Health*. 1990;44(4):257–258. doi: <https://doi.org/10.1136/jech.44.4.257>
159. Greenland S, Neutra R. Control of confounding in the assessment of medical technology. *Int J Epidemiol*. 1980;9(4):361–367. doi: <https://doi.org/10.1093/ije/9.4.361>
160. Robins JM. Data, design, and background knowledge in etiologic inference. *Epidemiology*. 2001;12(3):313–320. doi: <https://doi.org/10.1097/00001648-200105000-00011>
161. Sagiv SK, Tolbert PE, Altshul LM, Korrick SA. Organochlorine exposures during pregnancy and infant size at birth. *Epidemiology*. 2007;18(1):120–129. doi: <https://doi.org/10.1097/01.ede.0000249769.15001.7c>
162. World Health Organization. *Body Mass Index (BMI)*. 2007. Available online: http://www.euro.who.int/nutrition/20030507_1. Accessed on September 10, 2007.
163. Beral V. Breast cancer and hormone-replacement therapy in the Million Women Study. *Lancet*. 2003;362(9382):419–427. doi: [https://doi.org/10.1016/s0140-6736\(03\)14065-2](https://doi.org/10.1016/s0140-6736(03)14065-2)
164. Hill AB. The environment and disease: Association or causation? *Proc R Soc Med*. 1965;58(5):295–300. doi: <https://doi.org/10.1177/003591576505800503>
165. Vineis P. Causality in epidemiology. *Soz Präventivmed*. 2003;48(2):80–87. doi: <https://doi.org/10.1007/s00038-003-1029-7>
166. Empana JP, Ducimetiere P, Arveiler D, et al. Are the Framingham and PROCAM coronary heart disease risk functions applicable to different European populations? The PRIME Study. *Eur Heart J*. 2003;24(21):1903–1911. doi: <https://doi.org/10.1016/j.ehj.2003.09.002>
167. Tunstall-Pedoe H, Kuulasmaa K, Mahonen M, et al. Contribution of trends in survival and coronary-event rates to changes in coronary heart disease mortality: 10-year results from 37 WHO MONICA project populations. Monitoring trends and determinants in cardiovascular disease. *Lancet*. 1999;353(9164):1547–1557. doi: [https://doi.org/10.1016/S0140-6736\(99\)04021-0](https://doi.org/10.1016/S0140-6736(99)04021-0)
168. Cambien F, Chretien JM, Ducimetiere P, et al. Is the relationship between blood pressure and cardiovascular risk dependent on body mass index? *Am J Epidemiol*. 1985;122(3):434–442. doi: <https://doi.org/10.1093/oxfordjournals.aje.a114124>
169. Hosmer DW, Taber S, Lemeshow S. The importance of assessing the fit of logistic regression models: a case study. *Am*

- J Public Health*. 1991;81(12):1630–1635. doi: <https://doi.org/10.2105/AJPH.81.12.1630>
170. Tibshirani R. A plain man's guide to the proportional hazards model. *Clin Invest Med*. 1982;5(1):63–68.
171. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. *Am J Public Health*. 1998;88(1):15–19. doi: <https://doi.org/10.2105/AJPH.88.1.15>
172. Uter W, Pfahlerberg A. The application of methods to quantify attributable risk in medical practice. *Stat Methods Med Res*. 2001;10(3):231–237. doi: <https://doi.org/10.1177/096228020101000305>
173. Schwartz LM, Woloshin S, Dvorin EL, Welch HG. Ratio measures in leading medical journals: structured review of accessibility of underlying absolute risks. *BMJ*. 2006;333(7581):1248. doi: <https://doi.org/10.1136/bmj.38985.564317.7C>
174. Nakayama T, Zaman MM, Tanaka H. Reporting of attributable and relative risks, 1966–97. *Lancet*. 1998. Vol. 351, N 9110. P. 1179. doi: [https://doi.org/10.1016/S0140-6736\(05\)79123-6](https://doi.org/10.1016/S0140-6736(05)79123-6)
175. Cornfield J. A method of estimating comparative rates from clinical data; applications to cancer of the lung, breast, and cervix. *J Natl Cancer Inst*. 1951;11(6):1269–1275.
176. Pearce N. What does the odds ratio estimate in a case-control study? *Int J Epidemiol*. 1993;22(6):1189–1192. doi: <https://doi.org/10.1093/ije/22.6.1189>
177. Rothman KJ, Greenland S, Rothman KJ, Greenland S. Measures of Disease Frequency. In: *Modern epidemiology*. Rothman KJ, Greenland S, eds. 2nd ed. Lippincott Raven; 1998. pp. 44–45.
178. Doll R, Hill AB. The mortality of doctors in relation to their smoking habits: a preliminary report. *BMJ*. 1954;1(4877):1451–1455. doi: <https://doi.org/10.1136/bmj.1.4877.1451>
179. Ezzati M, Lopez AD. Estimates of global mortality attributable to smoking in 2000. *Lancet*. 2003;362(9387):847–852. doi: [https://doi.org/10.1016/S0140-6736\(03\)14338-3](https://doi.org/10.1016/S0140-6736(03)14338-3)
180. Greenland S. Applications of Stratified Analysis Methods. In: *Modern epidemiology*. Rothman KJ, Greenland S, eds. 2nd ed. Lippincott Raven, 1998. pp. 295–297.
181. Rose G. Sick individuals and sick populations. *Int J Epidemiol*. 2001;30(3):427–432. discussion. doi: <https://doi.org/10.1093/ije/30.3.427>
182. Vandenbroucke JP, Koster T, Briet E, et al. Increased risk of venous thrombosis in oral-contraceptive users who are carriers of factor V Leiden mutation. *Lancet*. 1994;344(8935):1453–1457. doi: [https://doi.org/10.1016/S0140-6736\(94\)90286-0](https://doi.org/10.1016/S0140-6736(94)90286-0)
183. Botto LD, Khoury MJ. Commentary: facing the challenge of gene-environment interaction: the two-by-four table and beyond. *Am J Epidemiol*. 2001;153(10):1016–1020. doi: <https://doi.org/10.1093/aje/153.10.1016>
184. Wei L, MacDonald TM, Walker BR. Taking glucocorticoids by prescription is associated with subsequent cardiovascular disease. *Ann Intern Med*. 2004;141(10):764–770. doi: <https://doi.org/10.7326/0003-4819-141-10-200411160-00007>
185. Martinelli I, Taioli E, Battaglioli T, et al. Risk of venous thromboembolism after air travel: interaction with thrombophilia and oral contraceptives. *Arch Intern Med*. 2003;163(22):2771–2774. doi: <https://doi.org/10.1001/archinte.163.22.2771>
186. Kyzas PA, Loizou KT, Ioannidis JP. Selective reporting biases in cancer prognostic factor studies. *J Natl Cancer Inst*. 2005;97(14):1043–1055. doi: <https://doi.org/10.1093/jnci/dji184>
187. Rothman KJ, Greenland S, Walker AM. Concepts of interaction. *Am J Epidemiol*. 1980;112(4):467–470. doi: <https://doi.org/10.1093/oxfordjournals.aje.a113015>
188. Saracci R. Interaction and synergism. *Am J Epidemiol*. 1980;112(4):465–466. doi: <https://doi.org/10.1093/oxfordjournals.aje.a113014>
189. Rothman KJ. *Epidemiology. An introduction*. Oxford: Oxford University Press; 2002. pp. 168–180.
190. Rothman KJ. *Interactions Between Causes. Modern epidemiology*. Boston: Little Brown; 1986. pp. 311–326.
191. Hess DR. How to write an effective discussion. *Respir Care*. 2004;49(10):1238–1241.
192. Horton R. The hidden research paper. *JAMA*. 2002;287(21):2775–2778. doi: <https://doi.org/10.1001/jama.287.21.2775>
193. Horton R. The rhetoric of research. *BMJ*. 1995;310(6985):985–987. doi: <https://doi.org/10.1136/bmj.310.6985.985>
194. Docherty M, Smith R. The case for structuring the discussion of scientific papers. *BMJ*. 1999;318(7193):1224–1225. doi: <https://doi.org/10.1136/bmj.318.7193.1224>
195. Perneger TV, Hudelson PM. Writing a research article: advice to beginners. *Int J Qual Health Care*. 2004;16(3):191–192. doi: <https://doi.org/10.1093/intqhc/mzh053>
196. *Annals of Internal Medicine. Information for authors*. Available online: http://www.annals.org/shared/author_info.html. Accessed on September 10, 2007.
197. Maldonado G, Poole C. More research is needed. *Ann Epidemiol*. 1999;9(1):17–18. doi: [https://doi.org/10.1016/S1047-2797\(98\)00050-7](https://doi.org/10.1016/S1047-2797(98)00050-7)
198. Phillips CV. The economics of 'more research is needed'. *Int J Epidemiol*. 2001;30(4):771–776. doi: <https://doi.org/10.1093/ije/30.4.771>
199. Winkleby MA, Kraemer HC, Ahn DK, Varady AN. Ethnic and socioeconomic differences in cardiovascular disease risk factors: findings for women from the Third National Health and Nutrition Examination Survey, 1988–1994. *JAMA*. 1998;280(4):356–362. doi: <https://doi.org/10.1001/jama.280.4.356>
200. Galuska DA, Will JC, Serdula MK, Ford ES. Are health care professionals advising obese patients to lose weight? *JAMA*. 1999;282(16):1576–1578. doi: <https://doi.org/10.1001/jama.282.16.1576>
201. Spearman C. The proof and measurement of association between two things. *Am J Psychol*. 1904;15(1):72–101. doi: <https://doi.org/10.2307/1412159>
202. Fuller WA, Hidiogrou M.A. Regression estimates after correcting for attenuation. *J Am Stat Assoc*. 1978;73(361):99–104. doi: <https://doi.org/10.1080/01621459.1978.10480011>
203. MacMahon S, Peto R, Cutler J, et al. Blood pressure, stroke, and coronary heart disease. Part 1, Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet*. 1990;335(8692):765–774. doi: [https://doi.org/10.1016/0140-6736\(90\)90878-9](https://doi.org/10.1016/0140-6736(90)90878-9)
204. Phillips AN, Smith GD. How independent are “independent” effects? Relative risk estimation when correlated exposures are measured imprecisely. *J Clin Epidemiol*. 1991;44(11):1223–1231. doi: [https://doi.org/10.1016/0895-4356\(91\)90155-3](https://doi.org/10.1016/0895-4356(91)90155-3)
205. Phillips AN, Smith GD. Bias in relative odds estimation owing to imprecise measurement of correlated exposures. *Stat Med*. 1992;11(7):953–961. doi: <https://doi.org/10.1002/sim.4780110712>
206. Greenland S. The effect of misclassification in the presence of covariates. *Am J Epidemiol*. 1980;112(4):564–569. doi: <https://doi.org/10.1093/oxfordjournals.aje.a113025>
207. Poole C, Peters U, Il'yasova D, Arab L. Commentary: This study failed? *Int J Epidemiol*. 2003;32(4):534–535. doi: <https://doi.org/10.1093/ije/dyg197>
208. Kaufman JS, Cooper RS, McGee DL. Socioeconomic status and health in blacks and whites: the problem of residual confounding and the resiliency of race. *Epidemiology*. 1997;8(6):621–628. doi: <https://doi.org/10.1097/00001648-199710000-00002>
209. Greenland S. Randomization, statistics, and causal inference. *Epidemiology*. 1990;1(6):421–429. doi: <https://doi.org/10.1097/00001648-199011000-00003>
210. Taubes G. Epidemiology faces its limits. *Science*. 1995;269(5221):164–169. doi: <https://doi.org/10.1126/science.7618077>
211. Temple R. Meta-analysis and epidemiologic studies in drug development and postmarketing surveillance. *JAMA*. 1999;281(9):841–844. doi: <https://doi.org/10.1001/jama.281.9.841>
212. Greenberg RS, Shuster JL. Epidemiology of cancer in children. *Epidemiol Rev*. 1985;7:22–48. doi: <https://doi.org/10.1093/oxfordjournals.epirev.a036284>
213. Kushi LH, Mink PJ, Folsom AR, et al. Prospective study of diet and ovarian cancer. *Am J Epidemiol*. 1999;149(1):21–31. doi: <https://doi.org/10.1093/oxfordjournals.aje.a009723>
214. Kemmeren JM, Algra A, Meijers JC, et al. Effect of second- and third-generation oral contraceptives on the protein C system in the absence or presence of the factor V Leiden mutation: a randomized trial. *Blood*. 2004;103(3):927–933. doi: <https://doi.org/10.1182/blood-2003-04-1285>
215. Egger M, May M, Chene G, et al. Prognosis of HIV-1-infected patients starting highly active antiretroviral therapy: a collaborative analysis of prospective studies. *Lancet*. 2002;360(9327):119–129. doi: [https://doi.org/10.1016/S0140-6736\(02\)09411-4](https://doi.org/10.1016/S0140-6736(02)09411-4)

216. Campbell DT. Factors relevant to the validity of experiments in social settings. *Psychol Bull.* 1957;54(4):297–312. doi: <https://doi.org/10.1037/h0040950>
217. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med.* 1999;130(6):515–524. doi: <https://doi.org/10.7326/0003-4819-130-6-199903160-00016>
218. Krinsky S, Rothenberg LS. Conflict of interest policies in science and medical journals: editorial practices and author disclosures. *Sci Eng Ethics.* 2001;7(2):205–218. doi: <https://doi.org/10.1007/s11948-001-0041-7>
219. Bekelman JE, Li Y, Gross C. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. *JAMA.* 2003;289(4):454–465. doi: <https://doi.org/10.1001/jama.289.4.454>
220. Davidson RA. Source of funding and outcome of clinical trials. *J Gen Intern Med.* 1986;1(3):155–158. doi: <https://doi.org/10.1007/BF02602327>
221. Stelfox HT, Chua G, O'Rourke K, Detsky AS. Conflict of interest in the debate over calcium-channel antagonists. *N Engl J Med.* 1998;338(2):101–106. doi: <https://doi.org/10.1056/NEJM199801083380206>
222. Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ.* 2003;326(7400):1167–1170. doi: <https://doi.org/10.1136/bmj.326.7400.1167>
223. 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(7):921–928. doi: <https://doi.org/10.1001/jama.290.7.921>
224. Barnes DE, Bero LA. Why re s on the health effects of passive smoking reach different conclusions. *JAMA.* 1998;279(19):1566–1570. doi: <https://doi.org/10.1001/jama.279.19.1566>
225. Barnes DE, Bero LA. Industry-funded research and conflict of interest: an analysis of research sponsored by the tobacco industry through the Center for Indoor Air Research. *J Health Polit Policy Law.* 1996;21(3):515–542. doi: <https://doi.org/10.1215/03616878-21-3-515>
226. Glantz SA, Barnes DE, Bero L, et al. Looking through a keyhole at the tobacco industry. The Brown and Williamson documents. *JAMA.* 1995;274(3):219–224. doi: <https://doi.org/10.1001/jama.1995.03530030039032>
227. Huss A, Egger M, Hug K, et al. Source of funding and results of studies of health effects of mobile phone use: systematic review of experimental studies. *Environ Health Perspect.* 2007;115(1):1–4. doi: <https://doi.org/10.1289/ehp.9149>
228. Safer DJ. Design and reporting modifications in industry-sponsored comparative psychopharmacology trials. *J Nerv Ment Dis.* 2002;190(9):583–592. doi: <https://doi.org/10.1097/00005053-200209000-00002>
229. Aspinall RL, Goodman NW. Denial of effective treatment and poor quality of clinical information in placebo controlled trials of ondansetron for postoperative nausea and vomiting: a review of published trials. *BMJ.* 1995;311(7009):844–846. doi: <https://doi.org/10.1136/bmj.311.7009.844>
230. Chan AW, Hrobjartsson A, Haahr MT, et al. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA.* 2004;291(20):2457–2465. doi: <https://doi.org/10.1001/jama.291.20.2457>
231. Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine-selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ.* 2003;326(7400):1171–1173. doi: <https://doi.org/10.1136/bmj.326.7400.1171>
232. Scherer RW, Langenberg P, von Elm E. Full publication of results initially presented in abstracts. *Cochrane Database Syst Rev.* 2007;(2):MR000005. doi: [10.1002/14651858.MR000005.pub3](https://doi.org/10.1002/14651858.MR000005.pub3)
233. 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(9263):1191–1194. doi: [https://doi.org/10.1016/S0140-6736\(00\)04337-3](https://doi.org/10.1016/S0140-6736(00)04337-3)
234. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA.* 2000;283(15):2008–2012. doi: <https://doi.org/10.1001/jama.283.15.2008>
235. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med.* 2001;134(8):663–694. doi: <https://doi.org/10.7326/0003-4819-134-8-200104170-00012>
236. Moher D. CONSORT: an evolving tool to help improve the quality of reports of randomized controlled trials. Consolidated Standards of Reporting Trials. *JAMA.* 1998;279(18):1489–1491. doi: <https://doi.org/10.1001/jama.279.18.1489>
237. Begg C., Cho M., Eastwood S., et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA.* 1996;276(8):637–639. doi: <https://doi.org/10.1001/jama.276.8.637>