

МЕТОДОЛОГИЯ
НАУЧНОГО ПОИСКА

УДК 167:61:573.01

КРИТЕРИИ ПРИЧИННОСТИ В МЕДИКО-БИОЛОГИЧЕСКИХ
ДИСЦИПЛИНАХ: ИСТОРИЯ, СУЩНОСТЬ И РАДИАЦИОННЫЙ АСПЕКТ.
СООБЩЕНИЕ 3. ЧАСТЬ 1: ПЕРВЫЕ ПЯТЬ КРИТЕРИЕВ ХИЛЛА:
ИСПОЛЬЗОВАНИЕ И ОГРАНИЧЕНИЯ#

© 2021 г. А. Н. Котеров^{1,*}

¹ Федеральный медицинский биофизический центр им. А.И. Бурназяна ФМБА России, Москва, Россия

*E-mail: govorilga@inbox.ru

Поступила в редакцию 18.10.2020 г.

После доработки 23.12.2020 г.

Принята к публикации 24.02.2021 г.

Сообщение 3 посвящено истории, сути и ограничениям эпидемиологических критериев причинности (“критерии Хилла”). На материале из оригинальных публикаций ведущих исследователей причинности (А.В. Hill., M.W. Susser, K. Rothman и др.; 1950-е гг. – 2019 г.), из десятков современных пособий по эпидемиологии и канцерогенезу, из документов международных организаций (НКДАР ООН, BEIR, USEPA, IARC и пр.), а также из множества прочих источников, в части 1 сообщения рассмотрены первые пять критериев Хилла: сила связи, постоянство ассоциации, специфичность, временная зависимость и биологический градиент (зависимость “доза–эффект”). Изложены теоретические и практические аспекты для каждого критерия: история появления, терминология, философская и эпидемиологическая суть, применимость в различных дисциплинах и ограничения. Для критерия сила связи представлены градации силы ассоциации (ординальные шкалы) по показателям относительного риска, отношения шансов и коэффициента корреляции Пирсона (Спирмена). Приведены фактические примеры в рамках каждого из критериев, включая данные из радиационной эпидемиологии и радиационной медицины.

Ключевые слова: критерии причинности, критерии Хилла, сила связи, постоянство ассоциации, специфичность, временная зависимость, биологический градиент

DOI: 10.31857/S0869803121030085

Состоящее из двух частей третье (предпоследнее) сообщение цикла посвящено сути и ограничениям критериев причинности. Если первые два сообщения [1, 2] могут быть названы вводными (второе – в том числе в историческом плане), то третье нацелено на конкретику и должно отразить все положения в рамках критериев, необходимые для методологии установления причинности эффектов от воздействий.

Наши сообщения выполнены на основе очень объемного материала: проанализированы сотни работ, посвященных проблеме причинности начиная с 1950-х годов, в том числе “классиков” этой темы, ведущих исследователей и аналитиков (почти все публикации – в оригиналах); равно как более 40 западных пособий по эпидемиологии, статистике в медико-биологических дисциплинах, а также по канцерогенезу, на многие сотни и даже тысячи страниц (“Oxford”, “Cambridge”, “Springer”, “Elsevier” и др.); целый ряд –

2018 и 2019 г.). Около 30 сходных западных пособий отражают проблему причинности для иных дисциплин – статистике и эпидемиологии (так) в экономике, социологии, юриспруденции и психологии. Также около 30 пособий по эпидемиологии, доказательной медицине и клиническим испытаниям являлись русскоязычными или переводными.

Использовано также множество документов международных организаций, связанных с радиационным фактором, канцерогенезом, токсикологией, окружающей средой и пр. В этих документах в правилах установления причинности связей обнаружено применение тех или иных (порой специальных) комплексов критериев причинности, но основа их все та же – “критерии Хилла” [3]. Десятки оригинальных методологий доказательности, используемых мировыми организациями (нередко предусматривающих “Weight of Evidence” или WoE – “Вес свидетельств” (перевод наш), понятие введено в 1960 г. [4]), включают, вновь, известные критерии при-

Публикуется в авторской редакции.

чинности [5]. Это связано с тем, что индуктивно-дедуктивные правила установления причинных зависимостей едины для человеческого разума и являются следствием законов логики, уходя корнями в построения философов прошлых столетий, преимущественно Д. Юма и Дж. Милля [1, 2].

Вероятно, можно утверждать, что представленный нами материал по теме причинности приближен к исчерпывающему как в плане охвата проблем, так и в плане наиболее современных представлений в этой области, причем не только для эпидемиологии, но и для иных описательных дисциплин (а также, как указывалось нами ранее, — и для обыденной жизни и деятельности [6]).

Два предыдущих сообщения в рамках цикла [1, 2], равно как представлявшиеся необходимыми пять наших публикаций-преамбул к сообщениям 3 и 4 (по критериям “Сила связи” (две работы [7, 8]), “Временная зависимость” [6], “Биологическое правдоподобие” [9] и “Эксперимент контрафактивный” [10]) детализированы настолько, насколько это было возможно (т.е. введен весь имевшийся на тему материал, даже косвенный).

Сообщение 3 выполнить в подобной форме реально только в виде монографии, что ныне не имеет смысла. Практически про каждый из девяти критериев причинности можно написать объемный обзор, как это и было отчасти сделано американским специалистом на стыке медико-биологических дисциплин, права, коммерции и политики Douglas L. Weed [11, 12] и нами [6–10]. Но вряд ли это целесообразно для большинства критериев.

Поэтому далее материал в двух частях Сообщение 3 представлен, по-возможности, в относительно сжатой форме, с максимальным использованием ссылок на наши предыдущие работы [1, 2, 6–10] (хотя информация с момента опубликования последних здесь несколько дополнена). Но тематический охват, как можно надеяться, от этого не пострадал. Из всех источников нами была извлечена и систематизирована *вся* имеющаяся там информация по теме (это следствие цифрового мира). Некоторые положения вынужденно повторяют в краткой форме то, что в предыдущих публикациях было изложено в более полном виде и с иллюстративными примерами.

Цель Сообщение 3 — развернутое изложение конкретного материала по всем девяти критериям Хилла во всех их аспектах. Вследствие его объема материал разделен на две части.

ОПРЕДЕЛЕНИЕ ПРИЧИННОСТИ В ЭКСПЕРИМЕНТАЛЬНЫХ И ОПИСАТЕЛЬНЫХ ДИСЦИПЛИНАХ [6]

В экспериментальных дисциплинах, где можно определять условия опыта и необходимые

контроли, получение доказательств причинности эффекта от воздействия достаточно ясно (экспериментальным называется подход, когда можно проконтролировать хотя бы один варьирующий фактор из многих [13]). Выявление *в эксперименте* статистически значимой ассоциации или корреляции — это *конечный* этап доказательности [14] (что относится и к рандомизированным контролируемым испытаниям в медицине — к RCT [13]). Но для описательных дисциплин, включающих вместе с эпидемиологией экономику, социологию, психологию, историю и др., доказательство ассоциации и выявление корреляции не означают каузацию (т.е. причинность) [3]. Какой бы статистической значимости ни была корреляция [1, 2].

Выявленная ассоциация в описательных исследованиях при отсутствии контроля над вариант(ами) может объясняться следующими неучтенными факторами (подробно почти все они разобраны в наших предыдущих публикациях [1, 2, 6]):

- случайностью (‘chance’) [6];
- вмешивающимися факторами (“третий фактор”; ‘confounder’; “конфаундер”) [1];
- систематическими ошибками (bias — “смещение”, “уклон”; десятки таковых перечислены, к примеру, в оксфордском словаре по эпидемиологии [15]);
- обратной причинностью (“протопатическое смещение”) и смежными с этим понятием конфаундингами показания, противопоказания и тяжести патологии/прогноза [6].

В связи с изложенным для эпидемиологии установление статистически значимой ассоциации между двумя явлениями, в отличие от экспериментальных дисциплин, только самый первый, *исходный* этап доказательства причинности [6] (точная цитата взята из [14]).

Чтобы подтвердить причинность ассоциации, вслед за постулатами Генле–Коха XIX века для инфекционных заболеваний [2], в 1950–1970-х гг. в эпидемиологии был разработан ряд “мер предосторожности” (“precautions”) — пунктов (“points”) — положений (“viewpoints”) — руководящих принципов (“guidelines”) — мнений (“judgments”) — критериев (“criteria”) — постулатов (“postulate”) и др. (см. в [1, 2, 6, 7]) для оценки причинности хронических, неинфекционных патологий. Названные термины означают по сути одно и то же, а качественные отличия их по силе утверждения служат только отражением позиций авторов. Наиболее известны девять критериев причинности Хилла (Austin Bradford Hill; 1897–1991; Англия) [3], восемь из которых этот авторитетный в прошлом английский статистик в области медицины только собрал воедино, взяв у других авторов [2]. Позже положения Хилла были развиты и дополнены М. Сассером (Mervin Wilfred Susser;

1921–2014; США) [16, 17] и А. Эвансом (Alfred Spring Evans; 1917–1996; США) [18].

Тем не менее теперь критерии причинности в эпидемиологии называются почти всегда “критериями Хилла” или “руководящими принципами Хилла”, включая объемные западные пособия по эпидемиологии, доказательной медицине и статистическим методам в иных дисциплинах. Далее мы тоже станем условно называть эти положения “критериями Хилла”, хотя историческое исследование показало, что такое название и неправомерно, и неэтично [2].

По сути, на что с самого начала указывал А.В. Hill, его “положения” (“viewpoints”) – это никакие не строгие правила-критерии, а некие принципы, руководства, чтобы оценить степень вероятности того, что ассоциация причинна [3]. Вслед за ним указанное построение повторяется практически в каждом соответствующем источнике на тему причинности. Ведь в философском плане получение абсолютного знания причинности для медико-биологических дисциплин – невозможно, даже в каком угодно эксперименте (см. в [1])¹ (список примечаний идет после основного текста).

Впрочем, один пункт или руководящий принцип является именно *критерием* и даже *постулатом*. Это – “Временная зависимость” (“Temporality”), порядок времени. То есть воздействие должно быть перед эффектом. Данный критерий подробно разобран нами ранее [6]. По философской логике он должен идти первым в списке критериев (как его и поставил M.W. Susser [17]), но исходно, и в списке Хилла [3], и в предшествующем ему комплексе из пяти положений в известном Сообщении Главного врача (Surgeon General) США о последствиях курения от 1964 г. [20], “Временная зависимость” стояла только четвертым пунктом (см. в [1, 2, 6]).

За прошедшие 50 с лишним лет комплекс критериев Хилла подвергался различным модификациям, обычно без указания на основания. И число критериев (первично девять, но бывает и много менее), и их порядок варьируют без какой-либо видимой системы [1], на что указывалось в специальном исследовании Weed D.L., Gorelic L.S., 1996 [21] (“В обзорах часто исключают и иногда изменяют критерии без объяснения причин таких изменений”; подробнее цитату см. в [1]). Наш беглый анализ показал, что за прошедшие 24 года, вероятно, ничего не изменилось, причем самодеятельность касается не только конкретных исследований, но и объемных обучающих пособий по эпидемиологии и документов международных организаций. Причины элиминаций и изменений никогда не называются, хотя почти всюду это – “критерии Хилла”. Попытки проанализировать и систематизировать основания для

указанных флуктуаций мало что могут прояснить, за исключением преобладания на первых местах некоторых основных критериев (сила ассоциации, постоянство ассоциации, зависимость “доза–эффект” и биологическое правдоподобие либо временная зависимость), как оно имело место и ранее, в источниках до 1996 г. [21].

Мы не станем вводить какие-то свои субъективные основания для порядка критериев, вопрос об их весе и приоритетах для доказательности не всегда однозначен, что намечено рассмотреть в Сообщении 4. Поэтому далее материал изложен в соответствии с первым появившимся полным списком критериев, т.е. из Hill A.V., 1965 [3]: сила связи; постоянство ассоциации; специфичность; временная зависимость; биологический градиент (зависимость “доза–эффект”); биологическое правдоподобие; согласованность с текущими фактами и теоретическими знаниями; эксперимент и аналогия. Первая часть Сообщения 3 охватывает первые пять критериев, а вторая часть – остальные четыре.

СИЛА СВЯЗИ (STRENGTH OF ASSOCIATION) [7, 8]

Подробно этот критерий рассмотрен в наших обзорах [7, 8].

Чем сильнее связь, тем меньше вероятность влияния непрчинных факторов, но слабая связь не отменяет возможность ассоциации [7]

Критерию “Сила связи [эффекта, ассоциации]” (“Strength of the association” [3]) придается большое значение при доказательствах причинности. Впервые он был введен одним из исследователей связи между курением и раком легкого, E.C. Hammond в 1955 г. [2, 7]; с тех пор многие авторы отводили ему одну из главных ролей [7].

Чем сильнее ассоциация, тем меньше вероятность влияния случайности, смещений (bias) и “третьих”, вмешивающихся факторов (конфаундеров) [7]. Об этом было написано еще Хиллом, который поставил данный критерий на первое место [3], а до него – исследователем эффекта курения, “классиком” силы связи (ниже) Jerome Cornfield (1912–1979; США) [7].

Указывается, что в случае конфаундеров более вероятно их влияние на слабые ассоциации, и менее вероятно, что они объясняют сильные эффекты [22]. В работе [23] предлагается методология определения того, что “наблюдения говорят сами за себя”: когда связь между воздействием и эффектом настолько “драматична” (dramatic), что смещения (bias) можно исключить заведомо. Собственно, еще А.В. Hill в своей программной публикации от 1965 г. [3] говорил, что в его практике (1920-е годы) бывали такие выраженные эф-

Таблица 1. Размер выборки (sample size), необходимой для получения статистически значимого RR разной величины в RCT или когортном исследовании ($p(\alpha) = 0,05$; мощность 80%, т.е. ошибка $\beta = 0.2$) [26]

Table 1. Sample size required to obtain statistically significant RR of different values in RCT or cohort study ($p(\alpha) = 0.05$; power 80%, ie error $\beta = 0.2$) [26]

RR	Число инцидентов сравнительно с контрольной группой			
	1/50.000	1/10.000	1/5000	1/1000
2	1.177.295	235.430	117.697	23.511
2.5	610.446	122.072	61.025	12.187
3	392.427	78.472	39.228	7.832
5	147.157	29.424	14.707	2.934
7.5	78.946	15.783	7.888	1.572
10	53.288	10.652	5.323	1.059

факты, когда и “статистики никакой не нужно”. В [23] приведена историческая сводка с подобными фактами при клинических испытаниях и предлагается ввести граничный коэффициент эффекта, равный 10 (“с высокой вероятностью отражает реальные эффекты, даже в присутствии конфаундеров”), как это имело место ранее для RCT в работе [24]. В то же время в Rawlins M., 2008 [25] отмечается, что названная граница величины отношения “сигнал/шум” основана на “впечатлении” и никак эмпирически не поддерживается.

В пособии по эпидемиологии издательства Springer от 2014 г. [26] приведена таблица расчета величины выборки (“sample size”), необходимой для получения значимых относительных рисков (RR) той или иной величины (табл. 1).

Критерий “Сила связи” не абсолютен: хоть и редко, но возможно наличие таких сильных конфаундеров, которые смогут имитировать сколь угодно высокий уровень связи [27, 28]². С другой стороны, слабая ассоциация не свидетельствует об отсутствии причинности, только в подобных случаях труднее найти/устранить третий фактор, смещение и пр. в качестве возможного объяснения эффекта; необходимы большие выборки и большее количество повторных изысканий. Для увеличения статистической мощности исследования в тех случаях, когда ассоциация слаба, служат синтетические подходы, объединяющие данные отдельных работ: мета-анализ и pooled-анализ [14]. Собственно, современная эпидемиология разных направлений (включая радиационную [7]) – это преимущественно слабые ассоциации, с малыми RR и OR (отношение шансов), ибо все сильные связи уже были открыты ранее [7, 14, 31–34]³.

Два основных показателя силы связи – RR/OR и r Пирсона [7, 8]

Для описательных дисциплин (эпидемиология, социология, экономика, юриспруденция,

история, психология и пр.) силу ассоциации нередко определяют по двум параметрам [7, 8]:

- по RR либо OR (в основном эпидемиология),
- по силе (иногда встречается русскоязычное: “теснота”) корреляции (редко и неправомерно – эпидемиология (объяснения ниже), но более – остальные перечисленные дисциплины).

В работах прежних десятилетий упоминается также об оценке силы связи по проценту атрибутивного риска для мультипричинных зависимостей [39, 40].

В отдельных источниках в качестве субкатегории критерия “Сила связи” называется другой критерий причинности Хилла – “Биологический градиент”, т.е. наличие дозовой зависимости (см. в [7]). Но это, вероятно, просто качественное свидетельство реальности самой связи, а не ее силы.

“Сила связи” входит в статистическую категорию “величина эффекта” (“effect size”) [41–43]⁴. Для величины эффекта в неэпидемиологических дисциплинах наиболее распространен показатель r – коэффициент линейной корреляции Пирсона [8].

Ординальные шкалы для диапазонов силы связи по RR и OR [7]

Ранее для эпидемиологии имелся как бы некий канон, введенный еще в 1950-х годах J. Cornfield, согласно которому слабая ассоциация начинается уже с $RR < 3.0$ (см. в [7])⁵. И так продолжалось еще порядка 30–50 лет с некоторыми вариациями, что рядом авторов, конечно, критиковалось (см. в [7])⁶. Затем общепринятая граница для “слабой” связи в весомых источниках снизилась до $RR < 2.0$ (1986–2016 гг.; в том числе работы упомянутого авторитета D.L. Weed от 1999–2004 гг.) [7]. Для радиационной эпидемиологии граница в $RR < 2.0–3.0$, вероятно, в 90% случаев элиминирует понятие о сильных эффектах облучения.

В большинстве публикаций при рассмотрении критериев Хилла, как сказано, имеется утверждение, что признак причинности — это “сильная связь”. Среди трех факторов, отвечающих за “Уровень достоверности доказательств” в доказательной медицине, первым называется “существенный размер эффекта” [46]. Но авторы, в том числе руководящих публикаций и пособий по эпидемиологии, весьма редко указывают, какую связь по RR считать “сильной” или “слабой”. В лучшем случае мелькают названные выше значения 2–3. Тем не менее поиск в разнообразных источниках (от материала университетов США on-line до пособий по эпидемиологии и канцерогенезу) показал, что градации для RR/OR все же имеются. Вот только их унификация — отсутствует. Нами был обнаружен ряд ординальных шкал для RR, как полных, так и отрывочных. Подробная сводка (42 источника) приведена в [7]; материал эклектичен и отвечает почти любой конъюнктуре (причем большинство источников — весомы). В русскоязычных публикациях ничего не найдено.

Оказалось, однако, что наиболее часто для RR ссылаются (1985–2015) [7] на “шкалу Монсона” (Richard R. Monson, США), из двух изданий эпидемиологии профессиональных воздействий (1980; 1990) [47]⁷. Нами был сделан вывод [7], что для RR наиболее целесообразно пользоваться шкалой Монсона, немного расширив ее в районе сверхсильных зависимостей градациями из иного источника:

- отсутствие эффекта (Ignorable): $RR = 0.9–1.2$;
- слабая связь (Weak): $RR = 1.2–1.5$, или $0.7–0.9$;
- умеренная связь (Moderate, Modest, Medium): $RR = 1.5–3.0$, или $0.4–0.7$;
- сильная связь (Strong): $RR = 3.0–10.0$, или $0.1–0.4$;
- очень сильная связь (Very strong, Infinite): $RR > 10.0$ [47]. Исходя из другого источника, хотя и менее весомого [48], величина RR до 40.0 считается ‘Dramatic’, поэтому мы предложили ограничить максимальный диапазон Монсона так: $RR = 10.0–40.0$, или $0.0–0.1$ (реципрокные границы — введены нами по логике; ср. с предыдущим пунктом). Это дало возможность добавить следующий диапазон.
- подавляющая связь (Overwhelming [48]): $RR > 40.0$ [48] (реципрокное значение придумайте сами, если это имеет смысл).

Отметим важность наименьшей категории приведенных канонов — отсутствие эффекта при $RR < 1.2$. То есть когда избыточный относительный риск (ERR) < 0.2 , то эффекта нет. Сразу же становятся отсутствующими все риски циркуляторных патологий при облучении в дозах $< 0.1–$

0.2 Гр, о которых утверждается, к примеру, в мета-анализах М.Р. Little [49] и в МКР3-118 [50].

В “быстром ответе” [51] на известное исследование E. Cardis и соавт. от 2005 г. о частоте рака в объединенной когорте работников ядерной индустрии 15 стран [52] указано, что выявленный эффект, который по показателю RR менее 1.2, вполне может быть не атрибутивен облучению, а объясняться конфаундерами. И что это является основной эпидемиологической доказательности [53].

Но не все авторы, приводившие градации силы связи по RR, включают диапазон с отсутствием эффекта. С другой стороны, для некоторых исследователей подобный “нулевой” диапазон простирается до $RR < 1.5$ (см. в [7]). Все же, на наш взгляд, целесообразно придерживаться наиболее распространенной шкалы Монсона.

Встает вопрос о совместимости градаций для RR и OR. Таковых для OR почти не обнаружено, за исключением весомой публикации Rosenthal J.A., 1996 [42], в которой, как и в монографии Cohen J., 1988 [41], приведены градации для многих показателей в рамках “effect size”. В [42] имеется и единственная шкала для OR, которую нам удалось найти в весомом источнике:

- слабая связь: $OR = 1.0–1.5$;
- умеренная связь: $OR = 1.5–2.5$;
- сильная связь: $OR = 2.5–4.0$;
- очень сильная связь: $OR > 4.0–10.0$.

Следует иметь в виду, что “шкала Розенталя” [42] введена не для эпидемиологии, а для психосоциологии, когда требования к причинности и доказательности снижены [43, 54, 55]⁸.

Показатель OR рассматривается как информативный также при определении связи между генотипом и предрасположенностью к тем или иным заболеваниям (бинарные генетические тесты). В исследовании Рубанович А.В., Хромов-Борисов Н.Н., 2013; 2014 [56, 57] было показано, что при $OR < 2.2$ маркер обладает низкой прогностической эффективностью при любых частотах в популяции как заболевания, так и маркера. В то же время маркер может быть хорошим классификатором (когда для его носителей вероятность развития заболевания мала), если $OR > 5.4$, но при условии, что его популяционная частота достаточно высока (> 0.3).

Примеры “очень сильных” и “подавляющих” связей в эпидемиологии [7]

Можно усомниться в практической значимости названных верхних градаций для RR. Но в эпидемиологии имеются примеры не только “очень сильных”, но “подавляющих” и даже как бы “бесконечных” рисков. Источники для собранных нами данных приведены в [7], а здесь мы

только перечислим эти ассоциации (см. также прим. 2):

- прием диэтилstilбестрола беременными (“от невынашивания”) – аденокарцинома влагалища у их дочерей ($OR \sim \infty$, поскольку в единственном исследовании, типа “случай–контроль”, в группе “случаев” риск достигал 88%, а в группе “контроль” – отсутствовал);
- прием талидомида беременными (“успокоительное”) – частота пороков развития у потомства (RR до 8180);
- вирусный гепатит В – гепатоцеллюлярная карцинома ($RR = 232$);
- работа трубочистом – рак мошонки ($RR = 200$);
- внешнее облучение детей – рак печени (RR до 152);
- средства влажной дезинфекции – патологии легкого (RR до 116);
- винилхлорид – ангиосаркома печени ($RR = 91$);
- курение – рак легкого (RR до 40–60);
- внешнее или внутреннее облучение детей – рак щитовидной железы (RR до 56 и более).

Величина RR в значительной степени зависит, понятно, от редкости заболевания, т.е. низкой величины его фонового уровня [14, 31].

Из подборки следует, что и радиационные эффекты могут предусматривать “подавляющие” RR , но очень редко, когда фоновые значения показателя, действительно, ничтожны. В целом же, что отмечалось выше, современная эпидемиология, включая радиационную, есть скорее “слабые”, максимум – “умеренные” ассоциации, какое ее направление ни возьми [13, 50]. Как подчеркивается в ряде источников, радиация – сравнительно слабый (weak) канцероген [58–60] и мутаген [58], особенно в сравнении со многими химическими агентами [58]. Если добавить сюда: “и, иногда, умеренный” (до $RR < 3.0$), то получится совсем правильно. Но канцероген – доказанный на всех уровнях биологической организации, и потому МАИР (Международное агентство по изучению рака; IARC) относит радиацию к первой группе опасности [61].

Юридический аспект градации величины RR [7]

Имеются данные о том, какую величину риска в судах США принимают за атрибутивную при установлении компенсации за вредность. Основы так называемого “правила Дауберта” (“Daubert ruling”, “Daubert standard”) заложены по прецеденту 1993 г., когда родители детей, рожденных с серьезными дефектами, J. Daubert и E. Schuller, подали в суд иск на фармакологическую компанию, утверждая, что к дефектам привело употреб-

ление беременными препарата Bendectin [62]. Суд рассмотрел все уместные данные из медико-биологических дисциплин. В рамках правил Дауберта ныне в судах ряда штатов США положительное решение выносится только при $RR > 2.0$. Значение два имеет тот смысл, что при его превышении теоретическая вероятность причинности эффекта более 50% [19, 62–68].

Этот юридический подход к оценке доказательности по RR был обоснован, судя по всему, Philip Cole (США) в 1990-х годах [63], в работе, посвященной критериям причинности на стыке эпидемиологии, социологии и юриспруденции. С другой стороны, названное правило эпидемиологами критикуется, особенно с тех позиций, что причинность целого ряда ассоциаций с $RR < 2.0$ все же общепризнана [64, 66, 68].

Для эпидемиологии определение силы связи по r Пирсона может быть неправомерным [8]

В ряде источников указывается на возможность определения силы ассоциации в эпидемиологии не только по RR/OR , но и по r Пирсона (приводим все известные нам ссылки [36, 69–72]). В публикации, посвященной критериям Хилла в молекулярной эпидемиологии (Fedak K.M. et al., 2015 [72]), авторы вообще договорились до следующего:

“Сегодня статистическая значимость, а не степень ассоциации, является общепринятым критерием для оценки силы наблюдаемой ассоциации и, следовательно, ее потенциальной причинности” (здесь и далее перевод цитат наш. – А.К.)⁹.

Если рассматривать “степень ассоциации” как RR , то, разумеется, все наоборот (см., например [43, 67, 73, 74]¹⁰ и другие источники [7]). Согласно [74], если представить себе масштабное исследование, которое выявило ассоциацию с $p < 0.001$ [для, вероятно, даже ошутимого значения r], но RR при этом составил 1.05, то речь может идти только о значимом открытии очень слабой связи. А вовсе не о большой силе ассоциации в эпидемиологическом плане.

Сходная ситуация и в молекулярной эпидемиологии, в частности, в области поиска генетических маркеров заболеваний. Так, высокая статистическая значимость ассоциации между патологией и генетическим маркером является необходимым, но не достаточным условием использования подобного маркера для предсказания предрасположенности к заболеванию [56, 57].

Определение силы связи по r – это прерогатива психосоциальных наук [41–43, 54, 55] (см. прим. 8, а также [8]). В критических публикациях Ferguson C.J., 2009; 2015 [43, 54] рассматривается некорректность притягивания к проблеме малой величины эффекта по r и R^2 (коэффициент детер-

Таблица 2. Результаты испытания вакцины против полиомиелита в 1955 г. По данным [43], в сокращении
Table 2. The results of the test of the polio vaccine in 1955. According to [43], in brief

Группа	Полиомиелит присутствует (число случаев)	Полиомиелит отсутствует (число случаев)	r и R^2
Реальное исследование			
Без вакцины	115	201.114	$r = 0.011$; $R^2 = 0.12\%$
С вакциной	33	200.712	
“Уточняющее” исследование			
Без вакцины	115	0	$r = 0.74$; $R^2 = 55\%$
С вакциной	33	82	

минации¹¹⁾ в психологии иллюстраций из медицины и эпидемиологии. Так, в [43] разбирается неправомерный пример из психологического обзора [75] про малую корреляцию при испытании вакцины против полиомиелита в 1955 г. Контрольная и опытная группы насчитывали каждая более 200.000 человек (табл. 2).

Казалось бы, корреляция между приемом вакцины и заболеваемостью полиомиелитом в группе из 200 000 человек совершенно ничтожна (верх табл. 2; $R^2 = 0.12\%$). Но видно, что в контрольной группе имелись 115 случаев полиомиелита, а в группе с вакциной – только 33. Относительный риск (точнее, обратный показатель – снижение частоты заболевания) весьма большой: $115/33 = 3.5$. По критериям силы эффекта это – сильная ассоциация (см. выше).

Далее С. J. Ferguson [43] делает предположение, что, возможно, среди порядка 200000 человек в группах подавляющее большинство в процессе исследования могли вообще не контактировать с возбудителем полиомиелита (это ведь не респираторная инфекция). Потому узнать для них эффект вакцины на деле невозможно. Приняв условно допущение, что с вирусом полиомиелита в каждой группе могли контактировать только заболевшие в контроле 115 человек, получаем как бы “настоящую” выборку. Из 200000 человек почти все – пустой балласт (не встречались с вирусом). Из нижней части табл. 2 видно, что корреляцию между введением вакцины и заболеваемостью полиомиелитом надо рассматривать в двух “контактировавших” группах по 115 человек. В результате коэффициент корреляции становится весьма большой ($r = 0.74$), и величина эффекта составляет даже “юридически” [19, 62–68] корректную долю, $R^2 = 55\%$ (связь затрагивает более половины дисперсии). Отмечается [43], что полученная корреляция с $r = 0.74$ вполне согласуется и

с большой величиной [снижения] относительного риска (составляющей 3.5).

В [43] приведено также множество других примеров из эпидемиологии и медицины со сравнительными оценками величины r для исходной (net) выборки и для выборки на основе изучаемой гипотезы (“hypothesis-relevant approaches”). Величины r нередко отличаются, как и в случае с эффектом вакцины от полиомиелита, на порядок-полтора.

Итак, согласно [43], эпидемиологические работы могут оперировать данными о многих тысячах участников, из которых, по факту, лишь небольшая часть имеет отношение к гипотезе исследования (те, кто действительно подвергается воздействию рассматриваемого фактора, или же имеет риск развития патологии при отсутствии вмешательства). С другой стороны, при оценке в области психологии нового метода лечения, к примеру, депрессии, была бы отобрана выборка только из тех индивидуумов, у которых эта депрессия есть (“целевая выборка”). То есть все из них имели бы отношение к гипотезе исследования [43]. Конечно, возникает мысль, что и в медицине и в эпидемиологии могут быть полные выборки, соответствующие целевой гипотезе, например, при РСТ препарата на больших ишемических болезнях сердца. Но в таком случае и RR, и r будут, вероятно, достаточно высоки.

Ординальные шкалы для величины эффекта по r Пирсона [8]

Собранный попутно соответствующий материал, как сказано, мало актуален для эпидемиологии. Тем не менее, в [8] нами были представлены 19 обнаруженных шкал для диапазонов r , предложенные за период 1925–2019 гг. Большинство градаций разработаны для психо-социальных дисциплин; реже для собственно статистики и пр. И все шкалы – *разные*, хотя бы в чем-то. Поэтому данная коллекция оставляет тягостное впечатление анархии и, вновь, подходит для почти любой конъюнктуры [8].

В [8] мы не давали рекомендаций, какую шкалу использовать. Но здесь позволим себе привести шкалу D.E. Hinkle с соавт. от 1979–2003 гг. (первое – пятое издания) [76] для корреляции Пирсона в психо-социальных дисциплинах, поскольку с эпидемиологической позиции она представляется наиболее уместной (оригиналы пособия недоступны; шкала реконструирована по пяти другим независимым источникам [8]):

- не принимаемая в расчет корреляция (“Little if any”; “Negligible”): 0–0.3;
- слабая корреляция: 0.3–0.5;
- умеренная корреляция: 0.5–0.7;
- высокая корреляция: 0.7–0.9;

– очень высокая корреляция: 0.9–1.0.

Высокая корреляция, таким образом, начинается с R^2 более 50%, что логично. Отметим, вновь, наличие “нулевой” шкалы, границы которой, по существу, элиминируют результаты большинства психосоциальных исследований [41–43, 54].

В некоторых источниках шкалу D.E. Hinkle и соавт., равно как и другие шкалы для r Пирсона, относили и к коэффициенту ранговой корреляции Спирмена [8].

Неабсолютность критерия

Выше и в других источниках [1, 2, 6, 7] нами уже отмечалось, что среди девяти “критериев Хилла” настоящий критерий только один – “Временная зависимость” (“Temporality”). Остальные “критерии” – это просто руководящие принципы для принятия решений, которые не абсолютны. То есть каждый из восьми критериев необязателен для установления причинности, а только желателен (“Ни необходимы, ни достаточны” [77]).

Такая ситуация и с критерием “Сила связи”. Сильная связь в описательных дисциплинах не является строгим доказательством причинности – возможно наличие конфаундеров, имитирующих причинную ассоциацию почти какой угодно силы [1, 29, 30] (см. прим. 2), в то время как слабые связи не отменяют причинность [14, 28, 30, 31, 33–36] (см. прим. 3). Кроме того, если исходить из “пирога” (“торта”) компонентной причинности К. Rothman (см. в [1]), не каждая составляющая причина имеет сильную связь с патологией, которую она вызывает; сила ассоциации подобной причины зависит от распространенности в популяции других факторов, т.е. причин [30].

Все же надо указать на то, что, согласно рассмотренной выше шкале Монсона, для $RR = 0.9–1.2$ ассоциация считается незначительной. То есть с формальных позиций при $ERR < 0.2$ критерий “Сила связи” работает именно как критерий.

ПОСТОЯНСТВО АССОЦИАЦИИ (CONSISTENCY OF THE ASSOCIATION)

Неудовлетворительность термина

Значения слова “consistency” разные: “последовательность; логичность; постоянство; согласованность”. Google-переводчик “знает” только “согласованность”. В течение более чем 50 лет наблюдается терминологическая путаница.

Впервые данный критерий был введен, как и предыдущий, E.C. Hammond в 1955 г. [2] Оригинальный источник 1955 г. (монография) нам недоступен, но данные из него были восстановлены по поздним обзорам, в которых термин представлен в последней форме – “consistency” (ссылки см. в [2]). Дальнейшие публикации по названно-

му критерию имелись, и реконструировать истоки стало возможно:

- Sartwell P.E., 1955 [2]: “повторение и наращивание” (“repetition and amplification”) независимых исследований.

- Wynder E.L., 1956; 1961 [2]: “частота рака/иных патологий в популяции должна согласовываться (“consistent”) с распределением причинного агента”. Это иной смысл.

- Yerushalmy J., Palmer C.E., 1959 [2]: “множественный подход (“multiple approach”), т.е. исследование влияния фактора на заболевание в независимых типах наблюдений”.

- Lilienfeld A.M., 1959 [2]: впервые появляется термин “consistency”, причем в смысле “согласованность”: “как связано/согласовано распределение в популяции патологии и причинного фактора” (повторение из Wynder E.L., 1956 [2]). Указано, что “наличие consistency наблюдений... сходно с воспроизведением (“replication”) экспериментов”.

- Sartwell P.E., 1960 [2]: сформулировано уже строго: “подтверждение ассоциации путем воспроизведения” (“confirmation of the association by replication”); исследования разными авторами на разных популяциях.

Можно видеть, что именно A.M. Lilienfeld в 1959 г. ненароком сконструировал связку между согласованностью наблюдений за распределением в популяции патологии и причинного фактора, с одной стороны, и повтором исследований разными авторами и на разных популяциях – с другой (сравнив с повтором экспериментов).

В 1963 г. R.A. Stallones (США), готовивший раздел о причинности в Сообщение Главного врача США от 1964 г. о последствиях курения [20], “вспомнил что-то, что где-то когда-то прочитал” [2], и вот – появился критерий “The consistency of the association”, означающий, что различные методы и подходы должны давать сходные заключения. Это был первый критерий по списку из пяти в 1964 г. [2, 20].

В 1965 г. A.V. Hill [3] сформулировал все девять критериев, среди которых вторым стояла “Consistency”: “Неоднократно ли наблюдалась ассоциация разными людьми, в разных местах, при различных обстоятельствах и в разное время?” Видно, что Хилл имел в виду значение слова “consistency” не “согласованность”, а скорее, “постоянство” [ассоциации]. Но восьмым пунктом у Хилла стояла “Coherence”: *согласованность* с текущими фактами и теоретическими знаниями. Слово “coherence” переводится как “согласованность”, или, что здесь хуже, как “связь, связность”.

В последующие 50 с лишним лет не все авторы смогли уяснить разницу в понятиях¹². “Как отли-

чить coherence от consistency или plausibility (правдоподобия)?” (Rothman K., 2012 [30]). Удалить из списка критериев “Consistency”, т.е. воспроизводимость ассоциации, никто, понятно, не мог [21], зато “Coherence” часто постигала эта участь (смысл пытались заменить критерием “Биологическое правдоподобие” [30, 33, 78, 79], что не совсем правомерно, см. также в части 2 настоящего сообщения). Еще в 1996 г. в упоминавшемся исследовании [21] обнаружили, что среди 14 обзоров по критериям причинности “Coherence” была названа всего в двух. Мы тоже имеем примеры многих пособий по эпидемиологии последних лет, где в списке “критериев Хилла” указанный пункт просто отсутствует [28, 32, 71, 78] (и др.). Нет его и в основных отечественных учебниках по этому предмету [28] (и др.; см. ссылки в [6–8])¹³.

Наш анализ распространенности “Coherence” среди критериев Хилла в тех исследованиях, где они использовались как метод доказательности (поиск в PubMed на “Hill’s criteria”; 35 публикаций за 2013–2019 гг.), тоже показал, что встречаемость критерия наименьшая (74%). Одна из причин этого, по-видимому, – путаница формального смысла “Coherence” с “Consistency”.

В популярном пособии по эпидемиологии (часто цитируют) Gordis L., 2014 (пятое издание) [78] в критериях Хилла вместо “Consistency” четвертым пунктом идет “Replication” (хотя простое воспроизведение не отражает всех смыслов критерия; см. ниже). Но затем под номером восемь стоит “Consistency”, однако в смысле “Coherence” Хилла – “Consistency with other knowledge” [78]. Сходные замены понятий есть и в иных источниках (ссылки не приводятся). Как у классика: “Коровьев против “Панаев” написал “Скабичевский”, а Бегемот против Скабичевского написал “Панаев””.

В российской медицинской “Системе оценки достоверности научных доказательств и убедительности рекомендаций” (2012) “Consistency” также означает “согласованность доказательств”, или “степень совпадения результатов различных исследований” [46].

Некоторые авторы пособий представляют термин “Consistency by repetition” (“Постоянство (или “Согласованность”) при повторении”) [82] и “Consistent replication” (“Последовательность” (или “Согласование”) при воспроизведении) [26]. Эти термины были приведены, вероятно, вслед за работами M. Susser (“Consistency on [upon] replication”) от 1977–2001 гг. [17, 83–86]. Как уже говорилось, подобные названия не отражают всех нюансов критерия, поэтому M. Susser ввел в 1991 г. для других смыслов термин “Survivability” (“жизнеспособность” [гипотезы] [17, 86]; обсуждение намечено в Сообщении 4).

Почти во всех русскоязычных источниках критерий “Consistency” назван по-разному:

- Перевод пособия по доказательной медицине Флетчер Р. и др., 1998: “Устойчивость” [связи] (ссылку см. в [1, 2, 6]).

- Пособие по эпидемиологии Власов В.В., 2006 [28]: “Подтверждаемость связи”.

- Перевод пособия по доказательной медицине Гринхальх Т., 2006 г. (ссылку см. в [6]): “Согласуются ли между собой связи, обнаруженные в разных исследованиях?”.

- Пособие по эпидемиологии Покровский В.И. и др., 2007 (ссылку см. в [2, 6, 7, 10]): “Убедительность ассоциации – подтверждение причины различными данными”.

- Перевод критериев Хилла из оксфордского словаря по эпидемиологии от 2009 г. [15]: “Устойчивость: ассоциация стабильна при воспроизведении результатов в исследованиях в других условиях и другими методами”. Однако в отдельной статье “Consistency” (одно из двух значений составляет как раз смысл критерия Хилла) этот термин переведен как “состоятельность результатов повторных исследований” (а в скобках указано – “постоянство”).

- Русскоязычная анкета по установлению причинности из организации “Врачи без границ” (2015) (ссылку см. в [6]): “Систематичность” (наблюдение последовательности [причина – следствие] во всех исследованиях...).

Мы со своего обзора 2010 г. [87] переводим критерий “Consistency” как “Постоянство ассоциации” [2] (и др.; см. в [2]). Судя по всему, при подготовке этого обзора (2009) [85], когда у нас не было русскоязычных источников, мы перевели “consistency” как “постоянство”, взяв наиболее подходящий по смыслу термин. Такой же перевод (“постоянство ассоциации”) есть в упомянутых (прим. 13) учебных пособиях Первого МГМУ им. И.М. Сеченова от 2011 г. [80] и КГМУ от 2017 г. [81]. Видимо, это параллельные озарения.

Итак, мы имеем “русскоязычный” критерий воспроизводимости эпидемиологических исследований в следующих словесных формах (в порядке весомости источников): устойчивость, подтверждаемость, состоятельность, постоянство, согласованность, убедительность, систематичность.

“Дурная бесконечность” и отсутствие даже намеков на попытки русскоязычной унификации (включая статьи в одном и том же энциклопедическом словаре [15]) показывают, что дела с использованием методологий причинности, известных уже более чем 50 лет, в отечественной науке не слишком хороши.

Наилучшие названия, вероятно, это “подтверждаемость” и “постоянство”. Исходя из нашей традиции – возьмем последнее.

Философская суть критерия: репликация данных – основа индуктивного подхода к доказательству причинности, но она не дает свидетельств согласно дедуктивной методологии К. Поппера

От ряда закономерностей – к обобщению, вот почему воспроизведение данных, в том числе при разных количественных и качественных условиях, является основанием для последующих индуктивных рассуждений о причинности [33, 85, 86]. Обыденный опыт показывает, что причинные эффекты, как правило, встречаются достаточно широко, в то время как ложные, случайные ассоциации локальны [71]. Однако, как отмечает М. Susser, критерий “Постоянство ассоциации” не согласуется с дедуктивной методологией К. Поппера, основанной на подтверждении и фальсификации гипотез. Конечно, если гипотеза не сможет пройти тесты на повторяемость, она окажется сфальсифицированной. Но никакое количество воспроизведений не докажет гипотезу; оно только продемонстрирует ее жизнеспособность. Таким образом, тест по критерию “Постоянство ассоциации” не подтверждает и не проверяет теорию, и не изменяет вероятность того, что она верна [85].

В свое время еще Д. Юм (D. Hume; 18 в.) указывал, что повторное наблюдение только влияет на нашу веру в то, что ассоциация является причинной. И в философском плане репликация зависимостей (индуктивная логика) не имеет смысла для подтверждения достоверности, поскольку ошибки могут повторяться [88] (равно как и конфаундеры или, например, неизвестная общая первопричина для предполагаемых “причины” и “следствия”) [89].

Тем не менее все мы знаем, что на практике и в описательных, и в экспериментальных дисциплинах индуктивные подходы неотъемлемы от доказательности. Например, в пособии от 2019 г. указано [33]: “Постоянство (consistency) результатов эпидемиологических исследований становится основой индуктивных рассуждений и высоко ценится эпидемиологами в качестве руководства для определения причинно-следственных связей... особенно при слабых ассоциациях”.

Индуктивные подходы – это то, чем мы руководствуемся в обыденной жизни. Стандартный пример [90, 91]: каждый раз мы ожидаем, что Солнце взойдет и завтра, раз оно ежедневно восходило ранее, но это – никак не гарантировано. Каждый рабочий день кто-то идет к метро или на остановку автобуса, ожидая, что и метро, и автобус будут функционировать, как и вчера (или вообще будут). Наконец, подбросив камень, мы

ожидаем, что он упадет на землю, хотя не имеем никаких данных ни о механизме сил притяжения, ни о том, всегда ли они действуют. Сходные с нашими рассуждения есть и у других авторов: “...кто считает, что мы используем индукцию каждый день, чтобы улаживать все аспекты нашего окружения” [92]¹⁴; “Индуктивные методы – это часть структуры нашей жизни, природы человека” [93]¹⁵; “Говорить о законности индукции бессмысленно, потому что индуктивные аргументы служат для определения того, что такое рациональные аргументы” [93]¹⁶; “Если мы действительно ограничимся дедуктивными рассуждениями, мы будем вынуждены отбросить большинство наших представлений о мире и отказаться от всякой надежды на знание природы, не говоря уже о надежде на улучшение здоровья людей” [94]¹⁷.

Эпидемиологическая суть критерия: воспроизведение данных, как аутентичное, так и при качественных и количественных модификациях, снижает вероятность посторонних вмешательств

“Один свидетель – не свидетель” (“testit unus testit nullus”). Это правило в том числе св. Инквизиции [95]¹⁸ распространяется и на эпидемиологию разных направлений, и на эксперимент:

“Одно исследование редко, если вообще когда-либо, является убедительным” (Susser M., 1986) [84]¹⁹. “Никто не должен верить единичному исследованию, поскольку в нем может быть допущена ошибка, не очевидная ни автору, ни интерпретаторам” (Pharmacoepidemiology, 2000) [96]²⁰. “Ни одно исследование не может дать единственного и окончательного решения проблемы” (Власов В.В., 2006 [28]).

В редакционной статье в “New England Journal of Medicine” от 1990 г. [22] разбирается вопрос о том, как должны реагировать практические медики, когда эпидемиологическое исследование выявило новый фактор риска. Указывается, что не следует советовать своим пациентам менять образ жизни *на основе только одного исследования*, даже хорошо выполненного и правдоподобного, *если риск невелик*. Целесообразно ожидать подтверждения другими авторами. “Возможностей для смещения столь много, что одно исследование редко бывает убедительным” [22]²¹.

Конечно, как и в случае правил Инквизиции [95], могут иметься исключения, так сказать, “в рабочем порядке”. На это указано еще в Hill A.B., 1965 [3] (“...будут случаи, когда повторение отсутствует или невозможно, и все же мы не должны стесняться делать выводы”²²), в Susser M., 1986 (“не каждая гипотеза дает возможность для повторения исследований”) [84]²³ и в других автори-

тетных источниках; можно вспомнить также “предупредительный принцип”, когда общественная значимость превалирует над научной достоверностью [1]. Неабсолютность критерия рассматривается далее, но правилом в медико-биологических дисциплинах является все же воспроизводимость:

“Состоятельность (постоянство) результатов повторных исследований — важный критерий в суждениях о причинности” (оксфордский словарь по эпидемиологии) [15]. “Самая мощная проверка” (Susser M., 1986) [85]²⁴. “Один из самых сильных аргументов в пользу причинности” (US Environmental Protection Agency, аббревиатуры US EPA или USEPA, 2005) [97]²⁵.

Исходный смысл этого критерия (Сообщение Главного врача США о последствиях курения от 1964 г. [20], Hill A.V., 1965 [3]), который был воспроизведен и дополнен в последующих работах иных авторов, следующий:

Вероятность искажения ассоциации случайностью [3, 20, 26, 32, 67, 97]²⁶, смещениями (bias) [26, 67] или вмешивающимися факторами [20, 29, 40, 67, 98] уменьшается, если имеется соответствие (“согласованность”) между данными, полученными:

- для разных выборок, страт, популяций [15, 28, 67, 69, 84, 85, 96, 97];
- в разных местах [69, 84, 85, 96, 99] (“географическая локализация или географическая обусловленность воздействия” [40])²⁷;
- в разное время [15, 69, 84, 85];
- при разных обстоятельствах [15, 28, 33, 69, 70, 84, 85, 100];
- различными методами [3, 15, 20] (при разных подходах (approach) [101] и дизайнах исследования [3, 67, 84, 96, 100]);
- разными исследователями [15, 67]²⁸.

Как отмечается в [102], репликация (воспроизведение) исследований служит универсальным решением при множественности эффекта (multiplicity) и анализе такового в подгруппах.

Особенно важным подкреплением причинности является получение однородных результатов разными методами, поскольку в работах с одинаковой методикой возможны одни и те же ошибки, смещения или конфаундеры [3, 102]. Необходимы разные исследования с неодинаковым дизайном, различными методами сбора данных и разным анализом результатов [102]. A.V. Hill указывал, что желательны совпадения, например, результатов проспективных и ретроспективных исследований [3], как и было в свое время с доказательствами эффектов курения [2, 20]. А в настоящее время, например, — с регистрацией повышенной частоты раков и лейкозов после компьютерной томографии в детском возрасте. Первая

такая публикация 2012 г. вызвала множество критических откликов; но затем были получены аналогичные данные по разным странам (более десятка исследований, “случай—контроль” и когортные) [6]. И хотя атрибутивность облучению здесь во многом сомнительна (обратная причинность) [6], тем не менее факт следования критерию “Постоянство ассоциации”, причем почти по всем перечисленным в списке выше пунктам, налицо.

*Сколько раз надо воспроизводить данные?
Это вопрос научного суждения*

В работе Susser M., 1986 [85] отмечается, что, хотя достижение статистической значимости и нацелено на определенное число повторов исследования, ответ на вопрос, когда фактическое наполнение критерия “Постоянство ассоциации” приведет к доказательству причинности, является прерогативой суждения, причем в качественном, а не в количественном смысле.

В пособии по предупреждению рака, в главе по причинности эффектов (Goodman K.J., Samet J.M., 2018 [36]) указано следующее:

“Нет никаких зафиксированных руководящих принципов по вопросу о том, сколько исследований должно быть сделано, насколько разнообразными они должны быть, и насколько [собранные данные] уместны, чтобы гарантировать успех [критерия] “Постоянство ассоциации”. Это — дело научного суждения; точные подходы предложить нельзя”²⁹.

Методология в рамках критерия: количественный подход (простая репликация — увеличение статистической мощности) и качественный подход (изменение условий для проверки гипотезы)

Повторение исследований и накопление однородных данных, помимо подтверждения самого факта ассоциации, увеличивают статистическую мощность исследования [69], что дает возможность объединять данные в мета- или pooled-анализах [14, 26, 33, 64, 67, 70] (и мн. др.). Однако, как отмечалось выше, простое воспроизведение результатов и повторение наблюдений (replication, repetition) не отражают всех нюансов критерия (которые можно видеть в списке в предыдущем подразделе). При простом воспроизведении, с одними и теми же дизайном и методиками, искажающие вмешательства (смещения, конфаундеры) тоже могут воспроизводиться [33, 36, 67, 103]³⁰. Только случайность сюда, ясно, не входит [36], хотя и она называется в подобном контексте в [67], вероятно, ошибочно.

В отличие от RCT или интервенционистских исследований, где, теоретически, можно устранить влияние не только случайностей, но и сме-

шений и вмешивающихся факторов, для описательных исследований мета- и pooled-анализы не способны доказать причинно-следственную связь статистическими методами. Нет гарантий того, что в объединяемых работах с отсутствующей рандомизацией не имелось смещений и/или неизвестных конфаундеров, поэтому для выводов необходимо учитывать другие, биологические и эпидемиологические соображения [9, 36, 67] (см. также в части 2 настоящего сообщения). Как отмечается в Weed D.L., 2000 [104], только мета-анализ, без иных подходов, не может использоваться для оценки причинно-следственной связи в эпидемиологии. На наш взгляд, сюда все же надо добавить поправку на силу связи. Вероятно, при высокой величине RR (см. выше) и значительном числе работ, мета- и pooled-анализы в большинстве случаев позволят сделать адекватные выводы, по крайней мере для практики.

М. Susser в своем комплексе критериев (именно такая категория) причинности первоначально ввел положение “Consistency on replication” (исходная монография 1973 г. [16] недоступна, но термин появляется по крайней мере в работе 1977 г. [83], встречаясь далее минимум до 2001 г. [17, 84–86]). Поняв, однако, что указанный термин не полностью отражает суть положения, М. Susser в 1991 г. ввел, как уже отмечалось, добавочный критерий “Survivability” (как субкласс “Consistency” [17]) – “Жизнеспособность” при проверке гипотезы различными тестами. Критерий отражает изменения дизайна исследований и “выживаемость” гипотезы при таких изменениях [17, 86].

В некоторых эпидемиологических дисциплинах критерий может пониматься шире, чем в классической эпидемиологии

Термин и понятие “классическая эпидемиология” встретились нам в ряде западных источников, включая объемное пособие [26] и словарь [15] по эпидемиологии.

Относительно неправомерного уподобления критерия “Consistency” другому критерию, “Coherence”, говорилось выше (см., к примеру, [30]).

Этот же критерий, т.е. “Постоянство ассоциации”, в молекулярной эпидемиологии порой толкуют близко к критерию “Биологическое правдоподобие”. Предложено рассматривать положение “постоянство” через интеграцию с экспериментами на молекулярном уровне и *in vitro*, которые должны послужить подтверждением гипотезы о механизмах эффекта. Цель данного подхода – уменьшить необходимость репликации собственно эпидемиологических исследований. Указывается, что расширение критерия “Consistency” до критерия “Biological plausibility” актуально для токсикологии, когда получение данных для людей затруднительно [72].

Сходным образом, критерий “Consistency” в рамках токсикологии (ВОЗ; Международная программа по химической безопасности³¹) расширяют до критерия “Аналогия” (“Analogy”) применительно к постоянству эффекта в группах структурно родственных химических агентов (“consistency across structurally related chemicals”) [105].

Вряд ли подобное размывание границ и смыслов критерия “Consistency” можно рассматривать всерьез, но, вновь, цитированные положения открывают возможности для многих желательных конъюнктур. В частности, для столь сомнительной методологии, как телеанализ (подробнее см. в [9]).

Неабсолютность критерия

Данные, раскрывающие аспекты проблемы, дробно представлены во многих источниках. Наиболее систематизировано и полно, с примерами из эпидемиологии и доказательной медицины, материал отражен в пособии от 2019 г. [33]. Ниже мы заменили большинство таких примеров на факты из области радиационной эпидемиологии.

Постоянство (воспроизводимость) связи не обязательно означает причинность.

Возможны следующие искажения:

1. Отражение публикационного смещения (положительные результаты чаще публикуются и авторами, и редакциями, чем отрицательные) [15, 33].

2. Тиражирование недоказанных и сомнительных результатов с “подкреплением” путем цитирования данных друг друга внутри узкого круга. Это наш пункт. Иллюстрацией служит пример с якобы свидетельствами о серьезных наследственных генетических эффектах облучения после аварии на Чернобыльской АЭС. Такие “свидетельства” получены в России и на Украине. Поскольку, согласно НКДАР, МКРЗ и BEIR, наследственные генетические эффекты у людей не обнаружены даже для масштабных когорт потомков облученных родителей (невозможно зарегистрировать ни в каких реальных исследованиях; об отдельных исключениях на субгеномном уровне см. в [9] и части 2 настоящего сообщения), то российским и украинским авторам, чтобы подкрепить свои изыскания (некорректные с эпидемиологических позиций), приходится ссылаться друг на друга или на маргинальные и старые западные работы, чтобы имитировать массовость (см. в [106–108]).

3. Намеренное искажение данных, включая повторение таких исследований. Множество подобных примеров приведено в монографии одного из лидеров скандинавской доказательной медицины Питера Гетше (P.C. Gotzsche) [109]. В работе Susser M., 1986 [84] также указывается, что исследователю, работающему в корпорации или отрасли, может быть предложено опубликовать

только те отчеты, которые не показывают связи между воздействием и неблагоприятным эффектом. Так, при изучении частоты опухолей головного мозга у работников нефтехимической промышленности возникла ситуация, когда две группы авторов, анализируя один и тот же набор случаев, пришли к противоположным выводам. В отчете независимых исследователей ассоциация была, а в докладах, представленных крупной химической корпорацией, таковая отсутствовала (см. в [84]).

4. Простая репликация работ может приводить и к такому же воспроизведению смещения либо конфаундера [33, 36, 67, 103] (см. также прим. 2 и 30).

Отсутствие постоянства связи не свидетельствует об отсутствии причинности.

Данный факт может быть обусловлен следующим:

1. Иной раз, как уже говорилось, повторение воздействия просто невозможно: условия уникальны, в особенности при авариях или экологических загрязнениях, но решения принимать надо (порой – в рамках “предупредительного принципа” [1]) [3, 71, 84, 110, 111].

2. Различиями в условиях воздействия (продолжительность, уровень), которые могут варьировать в разных популяциях [33, 103].

3. Различиями в сроках исследования применительно к латентному (инкубационному) периоду [33]³². Распределение случаев новообразований после воздействия фактора риска следует той же логарифмической зависимости, что и при инфекционных заболеваниях; одним из примеров этого является временная зависимость по выходу лейкозов в когорте пострадавших от атомных бомбардировок [112]. Когда минимальный латентный (или инкубационный) период еще не закончился, исследование недавней экспозиции агента не обнаружит ассоциации между ним и интересующим эффектом [33] (см. также в [6]).

4. Различиями в дизайне исследований и в аналитических стратегиях [32, 33, 96, 103, 111]. Ретроспективные и проспективные исследования, разные статистические модели, разная чувствительность и специфичность методов, а также статистическая мощность, продолжительность наблюдения и пр., все это может давать конечные отличия [33]. Примером влияния методологии на результат может служить изучение связи между диагностическим облучением головы и шеи (стоматология и др.) и учащением рака щитовидной железы. В ряде опросных исследований “случай–контроль” (реконструирующих по памяти посещения специалистов и рентгенодиагностику), был выявлен ошутимый эффект (OR до 2.7 [113] и до 3.5 [114]), включая зависимость от дозы [115]. Но замена опросов на анализ по медкартам, устраняющий “recall bias” (русскоязычный тер-

мин, к примеру, “смещение воспоминания” [28]), нивелировала разницу в частоте патологии между “случаями” и “контролями” [116].

5. Различиями в распределении компонентов причины в популяции [32, 33, 103, 111]. Неоднородность результатов может также отражать отличия в компонентных причинах созвездия достаточной причины (см. в [1]). Это лучше всего понимать в контексте модификации эффекта [33]. В Сообщении 1 нами приводились примеры компонентных достаточных причин радиогенного рака и циркуляторных патологий. Эффект могут модифицировать самые разные нелучевые причины (химикаты, курение, диета, стресс, возраст, ожирение, гипертония, мутации, дефекты апоптоза и репарации ДНК и др.) [1].

6. Различиями в стадии естественной истории процесса, лежащего в основе патологии [33]. Так, канцерогенез является многостадийным процессом, который нередко длится десятилетия [13, 26, 36]. Радиация способна не только инициировать его “с нуля”, но и потенцировать (вкуче с иными факторами), промотировать, а также модифицировать прогрессию уже имеющейся опухоли (“promotion/progression” [13]) на поздних стадиях развития [13, 117]. В связи с этим радиационное воздействие на индивидуумов с разными стадиями канцерогенеза (от его отсутствия до поздних этапов) может приводить к различиям не только в величине выявленных рисков, но даже в самом эффекте как таковом.

7. Различиями в эффективности вмешательств [33]. К примеру, использование вакцин в полевых условиях подчеркивает различие между эффективностью (efficacy) и действенностью (effectiveness). Потенциально эффективность вакцины может быть высокой, но если полевые условия не идеальны (ухудшения качества из-за отсутствия охлаждения или плохое восприятие целевой группой населения), ее действенность будет поставлена под угрозу [33] или окажется варьирующей от исследования к исследованию. Такая же ситуация может иметь место и при применении, скажем, противолучевых средств.

8. Различиями в изменчивости фактора риска [33]. Например, если диапазон воздействия фактора узок, то исследование может дать мало информации о его потенциальных последствиях для здоровья. Так, при использовании в качестве аналитических единиц индивидуумов не удалось найти постоянной ассоциации между потреблением соли и гипертонией, но в работах с экологическим дизайном, при сравнении на уровне стран (т.е. единиц с заметной межпопуляционной изменчивостью) корреляция появилась [33]. (Хотя экологические (корреляционные) исследования в эпидемиологии и не являются доказательными,

служа только для формирования гипотез [28, 78, 118, 119].)

СПЕЦИФИЧНОСТЬ АССОЦИАЦИИ (SPECIFICITY OF ASSOCIATION)

Критерий был предложен в публикации Yerushalmy J., Palmer C.E., 1959 (США; см. в [2]). По аналогии с постулатами Генле—Коха, авторы этой работы развивали “методологию и критерии” оценки причинности хронических патологий. Отсюда и появился пункт “Специфичность”, поскольку для инфекционных заболеваний, в целом, имеется строгое соответствие “возбудитель — патология” [2]³³. Иными словами, один фактор — один эффект.

Критерий при возможности его выполнения удобен для установления причинности. Но такие случаи для неинфекционных патологий очень редки

Неинфекционные патологии, причина которых специфична или относительно специфична (но не обязательно наоборот; см. ниже), действительно есть, хотя их и очень мало. Вот все примеры, которые удалось найти в массе источников (от нас добавлены два последствия облучения):

- бериллиоз как эффект бериллия [120];
- асбестоз [72] и мезотелиома легких [19, 26, 36, 70] после воздействия асбеста;
- ангиосаркома печени как результат экспозиции винилхлоридом [31, 36, 121];
- рак шейки матки после инфицирования вирусом папилломы [122];
- дети “цвета колы” от воздействия на беременных полихлорированными бифенилами (ПХБ) [84, 123];
- необычный спектр дефектов у новорожденных от воздействия на беременных талидомида (см. в [7]).
- средства влажной дезинфекции — специфичные патологии легкого [19, 124];
- общее облучение редкоизионизирующей радиацией в поглощенной дозе от 1 Гр — острый лучевой синдром (болезнь) человека [50];
- локальное радиационное воздействие на кожу в дозе 6–10 Гр излучения с низкой ЛПЭ — лучевой ожог (отличается от других типов ожогов) [50].

“Когда этот критерий выполняется, он обеспечивает чрезвычайно сильную поддержку выводу, что ассоциация является причинной” [96]³⁴. Но такие случаи, как видим, редки, хотя и можно встретить оптимистические утверждения типа: “большинство болезней имеют только конечное число причин” [99].

В 1950-х годах один из “диссидентов”, отрицавших причинную связь между курением и учащением рака легкого, Дж. Берксон (Joseph Berk-

son; 1899–1982; США), создатель модели регрессионного анализа и концепции “ошибки Берксона” при отборе групп (berkson bias), опирался именно на отсутствие специфичности эффекта курения (учащает множество патологий) [2]. Если верить пособию “Pharmacoepidemiology” от 2000 г. [96], то табачные концерны спустя много десятилетий все еще педалировали данный факт неспецифичности.

Специфичность эффекта, специфичность причины и специфичность восприимчивости

Критерий “Специфичность ассоциации” имеет разный вес при отнесении к эффектам и к причинам. Специфика ассоциации описывает точность, с которой появление одной переменной, за исключением других, будет предсказывать появление другой, также за исключением других [83, 85].

Специфичность эффекта данного фактора (в идеале, причина имеет только один эффект) [36, 70, 84, 85, 110, 111] мало подкрепляет причинно-следственную связь [84, 85, 110]. Так, у взрослых менингококк редко вызывает какую-либо патологию, кроме менингита. Но эта специфичность не дает менингококку никакого преимущества при установлении причины менингита, по сравнению, например, с гемофильной палочкой (*Haemophilus influenzae*), которая среди прочих патологий тоже вызывает это заболевание [84, 85].

Специфичность причины данного эффекта (в идеале, эффект имеет только одну причину) [36, 70, 84, 85, 110, 111], напротив, подкрепляет правдоподобие причинно-следственной связи [84, 85, 110] (см. список в предыдущем подразделе), причем даже для слабых ассоциаций [84, 110]. Однако с формальных позиций это ни необходимое, ни достаточное условие. Индуктивный подход, согласно К. Попперу, как уже говорилось, не дает окончательного доказательства. Отсюда следует, что если эффект, по всей видимости, пока объясняется только одной причиной, нельзя исключить, что иная причина просто еще не обнаружена [84, 85]. Надо сказать все же, что данные рассуждения М. Суссер [84, 85], основанные на отрицании К. Поппером доказательности индуктивных подходов, являются, на наш взгляд, в основном теоретическими³⁵. Поскольку, если вернуться к списку предыдущего подраздела, то станет ясно, что вряд ли мы будем учитывать на практике, что для лучевого синдрома может вдруг обнаружиться иная причина, чем облучение (не считая радиомиметиков), а для бериллиоза и мезотелиомы — иные причинные факторы, чем бериллий и асбест. Хотя оба последних агента вызывают и другие патологии [72, 125, 126] (нет специфичности эффекта), бериллиоз и мезотелиома индуцируются практически только ими (есть специфичность причины).

Специфичность восприимчивости. В отличие от предыдущих типов специфичности, этот феномен назван только в единственном источнике (Weiss N.S., 2002 [127]), с последующим пересказом в пособии 2011 г. [32]. Возможно, по сравнению с двумя предыдущими типами, это все же частный случай. Явление состоит в том, что связь между фактором и эффектом может наблюдаться только у индивидуумов, имеющих определенную характеристику, которая делает их восприимчивыми к воздействию фактора. Например, эритроциты лиц, дефицитных по глюкозо-6-фосфатдегидрогеназе, не могут реагировать на окислительный стресс путем формирования восстановленного глутатиона, что позволяет SH-группам окисляться. Прием сульфамидных препаратов, способных индуцировать реакции окисления, вызывает гемолитическую анемию, однако — только у лиц с дефицитом глюкозо-6-фосфатдегидрогеназы [127]. Сюда можно добавить пример с индивидуумами, обладающими различными дефектами в репарации ДНК, у которых воздействие ионизирующей радиации и ультрафиолета в дозах, переносимых здоровыми людьми, способно вызвать массу тяжелых последствий [128].

Специфичность эффекта для неинфекционных патологий определяют по силе ассоциации

Казалось бы, для неинфекционных патологий критерий специфичности применим только в единичных, исключительных случаях, примеры которых были представлены выше. Но если взять за основу вероятностный характер причины хронических патологий [1], и сравнивать относительные риски, то ситуация меняется.

В 1959 г., в том же номере журнала, в котором J. Yerushalmy и С.Е. Palmer предложили критерий специфичности [2], в работе Lilienfeld А.М., 1959 [2] была высказана его критика с последующим дополнением. Отмечалось, что, к примеру, курение способно вызывать множество патологий, но среди них именно рак легкого отличается наибольшим учащением, и, таким образом, вопрос специфичности все же решается: “специфичность эффекта должна интерпретироваться в терминах степени зависимости между фактором и заболеванием” (см. в [2]). Таким образом, именно А.М. Lilienfeld предложил оценку специфичности по степени RR. Затем этот подход появляется в Сообщении Главного врача США о последствиях курения от 1964 г. [20], равно как и в последующем таком же сообщении на тему от 2004 г. [69]. В нем указано: “В Сообщении [1964 г.] отмечен чрезвычайно высокий RR и атрибутивный риск для рака легкого у курящих, и сделано заключение, что ассоциация между курением и раком легкого имеет “высокую степень специфичности””³⁶. Но ссылок на работу Lilienfeld А.М., 1959 (см. в [2]) нет и в [20], и в [69].

В 1965 г. А.В. Hill в своей статье [3] также определил специфичность через силу ассоциации и привел пример с учащением рака легкого после курения³⁷. (Ссылка на других авторов, разработавших методологии оценки причинности, в работе [3], как известно, нет [2].) Следует отметить, что курение учащает рак и других тканей [20, 69], но специфичным “сайтом” фактора считаются именно легкие, в связи с несоизмеримо более высоким риском. МАИР распространяет данный тканевой подход и на иные канцерогены [100]. “Истинные зависимости специфичны, и при большинстве условий неправдоподобно, что препарат связан с увеличением развития множественных раковых образований” [129]³⁸. Добавим здесь: неправдоподобно, что в равной степени.

Кроме того, в Hill А.В., 1965 [3] есть рассуждения о возможной специфичности эффекта для той или иной группы работников, подвергавшихся воздействию конкретного химического вещества и пр. Вспоминаются “радиевые живописцы”, у которых величина ряда эффектов отражала специфичность воздействия α -излучателя ²²⁶Ra [117].

На основе названного подхода для многих причин неинфекционных патологий стало возможным устанавливать степень специфичности через силу ассоциации. В ряде источников приводятся примеры, хотя, конечно, некоторые из них выглядят несколько притянuto:

- Риск рака яичников в 3.2 раза выше у женщин с эндометриозом яичников сравнительно с генеральной популяцией [127]. N.S. Weiss [127] делает вывод, что связь рака яичников с эндометриозом яичников *специфична*, поскольку женщины с иными эндометриозами не имеют избыточного риска указанного типа рака.

- Рак полового члена у мужчин, согласно [130], *специфически* ассоциирован с увеличенной частотой смертности от рака шейки матки у их жен, так как связь с другими раками или причинами смерти отсутствует (вирус папилломы) [130].

- Ассоциация между применением снотворных/анксиолитических препаратов длительного действия пожилыми людьми и переломом шейки бедра *специфична*. Поскольку ассоциация не наблюдается для препаратов короткого действия с меньшими дневными седативными свойствами [127].

- Рак щитовидной железы может быть отнесен к *специфичным* для облучения патологиям, поскольку степень его учащения от радиационного фактора высока (НКДАР-2006) [131].

Отсутствие специфичности как негативный контроль при тестировании причинности

Дизайн “Негативный контроль” (“Negative Control”; см. в [132]) — это метод проверки наличия конфаундера для ассоциации эффекта с инте-

ресующим фактором. Подыскивается иное воздействие или эффект, которое может быть обусловлено тем же конфаундером, но ассоциация с которым лишена биологического правдоподобия. Если и здесь связь сохраняется, значит, налицо присутствие единого вмешивающегося фактора [132].

Есть примеры негативного контроля и на воздействие [132, 133], и на эффект [132, 134]:

- Курение беременных исходно рассматривалось как причина депрессии и гипертонии у потомства через прямой эффект *in utero*. Негативный контроль здесь – курение отца при беременности жены, когда нет существенного эффекта *in utero*, но конфаундер может быть тем же [132, 133].

- Курение дает одинаковую ассоциацию со смертностью как от суицидов (биологический механизм можно придумать), так и от убийств (правдоподобие отсутствует). В обоих случаях вероятно влияние поведенческих факторов [132, 134].

- Гормон-замещающая терапия в период постменопаузы в ранних исследованиях оказалась связанной со снижением смертности от сердечно-сосудистых патологий. Но при этом обнаружили такое же снижение смертности от убийств и суицидов, что невозможно объяснить правдоподобно (конфаундеры: образ жизни, поведение, социоэкономический статус и т.п.). RCT подтвердили это предположение (выявилась, напротив, положительная ассоциация со смертностью от сердечно-сосудистых патологий [26, 28, 99, 132]).

Подходы, основанные на критерии “Специфичность ассоциации”, служат для исключения артефактов как объяснения результатов работы [36, 77, 103, 127].

В обзоре [103] приведен следующий умозрительный пример. Когда реконструкция воздействия основана, скажем, на опросах и воспоминаниях, нет резона думать, что только некая конкретная врожденная аномалия будет обусловлена смещениями или конфаундерами: таковая тенденция должна иметь место со всеми аномалиями³⁹.

Есть и реальные примеры использования названного подхода. Так, было обнаружено, что при назначении циметидина (блокатора H₂-гистаминовых рецепторов) риск рака желудка увеличен приблизительно в 2 раза. Но то же самое имело место и при назначении антацидов (нейтрализаторов соляной кислоты). Хотя возможно, конечно, что само снижение уровня кислотности предрасполагает к развитию рака желудка, по мнению авторов оригинального исследования, отсутствие специфичности скорее подтверждает гипотезу, что еще недиагностированные опухоли или их предшественники побуждают пациентов искать терапию против раздражения желудка (см. в [127]).

Неабсолютность критерия

Еще Р.Е. Sartwell в 1960 г. [2] подверг критике руководящий принцип из Yerushalmy J., Palmer C.E., 1959 [2] “Специфичность ассоциации”, приводя пример множества патологий и последствий употребления этанола. А.В. Hill в 1965 г. также скептически отнесся к специфичности ассоциации на практике, хотя и поставил ее третьим критерием (“Я полагаю, что мультипричинность обычно более вероятна, чем одиночная каузация, хотя, возможно, если бы мы знали все ответы, то могли бы вернуться к единственному фактору” [3]⁴⁰).

С другой стороны, согласно мнению M. Susser [84], причинно-следственная связь, приближающаяся к зависимости один к одному, может оказаться искусственным построением. Артефакт возникнет из выявления эффектов какого-либо агента с последующей реклассификацией таких эффектов по отношению к нему задним числом. То есть для критерия “Специфичность ассоциации” может возникнуть проблема временного порядка: причина привела к критерию, или сам критерий создал причину?

Поскольку многие агенты вызывают рак в нескольких сайтах (тканях, органах), а многие типы рака имеют несколько причин, то критерий “Специфичность ассоциации” рассматривается как один из наиболее слабых возможных указаний на причинность (USEPA-2005 [97]). В ряде источников целесообразность критерия вообще отрицается. Вслед за К. Rothman и соавт. [30, 77], такой вывод сделан в пособиях по эпидемиологии [33, 78, 135] и в других работах [98]: “вводит в заблуждение” [30]; “не согласуется с современными знаниями” [33]; “недействителен как общее правило” [77]; “...является самым слабым из всех руководящих принципов и должен быть, по возможности, удален из списка” [78]; “можно безопасно исключить” [98]; “...считают это правило бесполезным” [135].

Некоторые авторы (в том числе пособий по эпидемиологии [28, 118, 136]) перешли от слов к делу и действительно устранили указанный пункт из списка критериев причинности [28] и даже из списка “критериев Хилла” [118, 136, 137] (на что А.В. Hill их явно не уполномочивал). В не раз упоминавшемся исследовании 1996 г., в котором анализировали частоту встречаемости того или иного критерия Хилла в обзорах, было обнаружено упоминание “Специфичности ассоциации” только в пяти источниках из 14 [21]. По нашей рассмотренной выше выборке работ, где использовались критерии Хилла (35 публикаций за 2013–2019 гг.), этот критерий встречался намного чаще (86%), но все равно находился только на седьмом месте из девяти.

Учитывая, что для руководящего принципа “Специфичность ассоциации”, как сказано, была предложена иная ипостась, базирующаяся на ве-

роятностной причинности и силе связи, причем ценность этого подхода обсуждалась во многих авторитетных источниках начиная с 1950-х годов [2, 3, 20, 69, 100], приведенная категоричность в отрицании похожа на массовый когнитивный диссонанс. Критерий, если рассматривать его в аспектах вероятности и величины RR, никакой не “самый слабый” и никакой не “бесполезный”.

Тем не менее формальный вывод один: специфичность ассоциации расширяет правдоподобие причинного вывода, но ее недостаток не отменяет его [78, 83, 103].

ВРЕМЕННАЯ ЗАВИСИМОСТЬ (TEMPORALITY) [6]

Подробно этот критерий, со многими примерами, рассмотрен в нашем обзоре [6].

Терминология [6]

Название критерия варьирует у разных авторов, начиная с 1950-х годов: обнаружено 18 англоязычных вариаций и восемь русскоязычных. Но в подавляющем числе случаев это “Temporality”, а на втором месте — “Temporal relationship”, причем обе конструкции есть в Hill A.V., 1965 [3]. Точным представляется термин M. Susser “Time order” [6, 17, 84, 85], но он почти не встречается у других авторов. Для русскоязычных названий наиболее удачным является, на наш взгляд, “Временная зависимость” (с ударением в первом слове; термин обнаружен только в переводе оксфордского словаря по эпидемиологии [15]) [6].

Философская суть критерия: воздействие должно быть перед эффектом [6]

Суть критерия, уходящая корнями в положение философов прежних столетий (в основном Д. Юма и Дж. Милля), проста и банальна: воздействие должно быть по времени перед эффектом. Несмотря на банальность, это положение повторяется во всех источниках по правилам причинности (см. в [6]).

Пункт “Временная зависимость” — единственный настоящий критерий или даже постулат в списке руководящих принципов причинности [6]

Исходя из факта, что время, согласно нашей практике, не течет вспять, руководящий принцип “Временная зависимость”, как уже говорилось, является единственным настоящим критерием (или даже постулатом) среди всех девяти пунктов Хилла. Он абсолютно необходим, и фальсификация по нему (согласно подходам К. Поппера) сразу же устраняет причинность, так что никакие другие критерии уже не нужны [6].

По логике данный критерий должен стоять в списке на первом месте, куда его и поставил

M. Susser (сразу после самого факта ассоциации), но в остальных источниках, включая Сообщение Главного врача США о последствиях курения от 1964 г. [20] и публикацию A.V. Hill от 1965 г. [3], ему уделена только четвертая позиция или даже далее (редко третья) [6] (нашу гипотезу о причинах этого см. в [6]).

Эпидемиологическая суть критерия: соблюдение латентного периода [6]

Появление патологии (или эффекта) должно регистрироваться спустя определенное время после воздействия, которое соответствует инкубационному периоду для инфекционных заболеваний, латентному периоду для неинфекционных заболеваний, либо периоду индукции эффекта. В большинстве источников под “латентным периодом” подразумевают время между причинным воздействием и обнаружением патологии (“эмпирический латентный период”). И только К. J. Rothman и соавт. считают, что термин “период индукции” должен использоваться для описания промежутка времени от причинного воздействия до необратимого возникновения заболевания, а термин “латентный период” — для промежутка времени от возникновения патологии до ее обнаружения [6].

Латентный период может быть уменьшен улучшенными методами диагностики. Период индукции, напротив, не может быть сокращен ранним выявлением заболевания, поскольку *появление болезни указывает именно на конец периода ее индукции*; болезнь не выявляется, пока период индукции не закончится. Хотя можно зарегистрировать как бы “предболезнь” по биомаркерам (“суррогатная конечная точка”; “Surrogate Endpoint”) [6].

Для солидного рака минимальный латентный период, согласно BEIR-VII, составляет по крайней мере пять лет [13]. Согласно же НКДАР-2000 — 8–10 лет после радиационного воздействия [138], хотя эффект канцерогенов обычно проявляется спустя десятилетия [71, 131]. Для лучевых лейкозов считается достаточным два года [13, 131].

Трудность соблюдения временного критерия, когда оба события в прошлом [6]

В случае экспериментального подхода, в том числе RCT, никаких сомнений в соблюдении критерия “временная зависимость”, понятно, быть не может. Иное дело для описательных дисциплин. Наилучшим образом критерий “Временная зависимость” соблюдается для когортных исследований, причем, ясно, проспективных. Наихудшем же — для методологий, когда оба события, предполагаемая причина и ее следствие, или произошли в прошлом (“случай—контроль”), или регистрируются в данный момент одновременно

(поперечные, “одномоментные” исследования), но причинно-следственные события все равно произошли ранее. Точная временная связь с помощью наблюдения здесь не может быть обнаружена и возможны разные временные смещения (см. ниже) [6].

Неприменимость критерия на практике в некоторых случаях [6]

Существуют такие события, для которых даже критерий “Временная зависимость” теряет смысл. Так, для генетической эпидемиологии, когда определяются ассоциации эффекта с сотнями тысяч генных вариантов, обычная каузальность не подходит. “Временная зависимость” неактуальна для врожденных генетических факторов как причины (из-за взаимодействия генов и окружающей среды – что здесь причина и следствие?) [6].

В случае хронических заболеваний время и характер начальных событий часто неясны, а между причиной и следствием могут существовать длительные скрытые периоды (субклинические фазы). Факторы риска, которые являются атрибутивными или предрасполагающими, могут изменяться со временем, что делает проблематичным определение временной зависимости их появления [6].

Наиболее же серьезным, часто скрытым, нарушением критерия “Временная зависимость” являются “Временные смещения” [6].

Временные смещения (“Temporal biases”) [6]

Объединяющий термин “Temporal bias” для эпидемиологии, хотя и представляется удачным, обнаружен только в единственном источнике – пособии Szklo M., Nieto F.J., 2019 (4-е издание [33]; термин встречается по крайней мере со второго издания от 2004 г., а первое недоступно). В эти смещения входят [6]:

Обратная причинность (“reverse causality” или “reverse causation”; другой синоним, введенный в 1980–1985 гг. Alvan R. Feinstein (США), – “протопатическое смещение”; “protopathic bias”). Феномен распространен не только в медицине и эпидемиологии, но и в социологии, психологии и в жизни (ниже). Суть множества определений состоит в том, что результат имеет обратный причинный эффект на воздействие. Для эпидемиологии – когда доклинические симптомы заболевания приводят к воздействию (а не наоборот), и за ним следует патология.

Конфаундинг показания (“confounding by indication”; понятие введено в 1980–1985 гг. O.S. Miettinen). Отражает тот факт, что терапия назначается лицам с определенным заболеванием, и, поэтому, существует сочетание эффектов терапии и заболевания на изучаемый результат. Конфаундинг

показания возникает, когда индивидуумы, которых лечат от их состояния, сравниваются с людьми, у которых нет ни состояния, ни лечения. Общий способ контроля за этим типом конфаундинга, помимо рандомизации, заключается в сравнении различных методов лечения для одного и того же состояния. Специфичность термина в том, что показанием к терапии должно быть *заболевание* как таковое, а не его *степень тяжести* или его *прогноз*. Данный конфаундинг крайне распространен при нерандомизированных испытаниях в фармакоэпидемиологии.

Конфаундинг противопоказания (“confounding by contraindication”; понятие введено, судя по всему, опять же O.S. Miettinen в 1983 г.). Суть явления в том, что противопоказания в исследуемой группе могут быть прогностическими критериями исхода побочных эффектов. Несмотря на аналогичную по форме между “confounding by indication” и “confounding by contraindication”, эти две проблемы не эквивалентны по значимости. Там, где есть показания к терапии, такие показания присутствуют регулярно и представляют собой проблему для неэкспериментальных исследований эффективности препаратов/воздействий. Напротив, какое-либо противопоказание выявляется только изредка, случайно, поэтому устранить подобный конфаундинг намного легче, просто элиминировав из выборки исключения.

Конфаундинг тяжести патологии/прогноза (“confounding by severity”, или “confounding by disease severity”, или “confounding by severity/prognosis”). Хотя конфаундинг тяжести патологии относятся к подтипу конфаундингов показания, тем не менее отличия ясны. В данном случае фактором является не заболевание/состояние, формирующее показание к воздействию, а его тяжесть (severity), которая служит основанием для иной, более интенсивной терапии и т.п. В результате может сформироваться представление, что более интенсивная и улучшенная терапия приводит к более тяжелым последствиям.

Способы контроля и устранения временного смещения [6]

Единого перечня методологий нигде не обнаружено, включая изданное “Springer” пособие по эпидемиологии от 2014 г. объемом порядка 2500 страниц [26]. Ниже представлена компиляция всей информации, которую удалось найти (ссылки см. в [6]):

1. Проведение, если это возможно, проспективных исследований, в которых точно устанавливается временная зависимость.

2. Проведение рандомизированных исследований. Важное значение придается Менделевской рандомизации (Mendelian Randomization), когда для исследований типа “случай–контроль” в качестве переменной для воздействия назнача-

ется генетическая детерминанта, не подверженная влиянию потенциальных конфаундеров окружающей среды и патологического состояния (исключает обратную причинность).

3. В перекрестных опросных исследованиях можно улучшить информацию о временной последовательности с помощью специальных вопросов.

4. Замена метода терапии для одного и того же состояния с последующим сравнением эффектов.

5. Исходная стратификация выборки, исключая лиц, имеющих до начала исследования интересующие последствия или их предпосылки.

6. Элиминация из выборки тех, у кого смерть зарегистрирована вскоре после постановки диагноза заболевания. То есть соблюдение правдоподобия латентного периода. По-видимому, это наиболее используемая методология.

Обратная причинность в жизни, социологии, психологии, в медицине и в эпидемиологии [6]

В работе [6] нами были собраны соответствующие факты из разных дисциплин и сфер жизни. Можно упомянуть связь между количеством огнестрельного оружия у населения и частотой убийств из него. По штатам США показана значимая прямая корреляция (хотя проспективного исследования никто не проводил). Но возможна обратная причинность: люди могут чаще приобретать огнестрельное оружие, когда они знают о более высоких показателях частоты убийств в своем месте проживания. И действительно, в регионах с большим распространением огнестрельного оружия регистрируется более высокая частота убийств и без огнестрельного оружия.

Некоторые примеры из медицины и эпидемиологии эффектны своей парадоксальностью (“перестал курить – увеличил частоту пневмоний и даже смертности от рака легкого” и т.п.). Значительное число случаев обратной причинности обнаружено для факторов риска сердечно-сосудистых патологий и рака [6].

Обратная причинность в радиационной эпидемиологии [6]

Все примеры лежат в рамках диагностических и терапевтических воздействий излучения (ссылки и более подробное рассмотрение см. в [6]):

Диагностические воздействия ^{131}I на щитовидную железу. В одном из исследований, если в оценку включались все пациенты, то RR для рака щитовидной железы был значимо выше (1.8), чем для генеральной популяции. Но когда пациентов разделили на две группы, “с подозрением на опухоль” и “по другим причинам”, то значимая ассоциация стала объясняться только первой группой (RR = 3.5). Для “других причин” RR = 0.9. Фено-

мен обратной причинности наблюдался в течение более чем 20 лет.

Терапевтические воздействия ^{131}I на щитовидную железу. Осуществляются по поводу незлокачественных патологий щитовидной железы (гипертиреоз, болезнь Грейвса, микседема и пр.) и приводят к учащению рака этого органа. Фактором, определяющим обратную причинность, являются сами указанные патологии, поскольку их наличие на порядки увеличивает вероятность рака щитовидной железы.

Воздействие Торотраста. Радиоактивный Торотраст (коллоид диоксида тория) ранее использовался для контрастного усиления при ангиографии черепа индивидуумов с подозрениями на опухоли мозга. У пациентов, которым проводили ангиографическое обследование вследствие подозрения на опухоль, был значимо повышен риск рака мозга независимо от того, включала ли процедура радиоактивный Торотраст или же нерадиоактивную контрастную среду. Вероятно, предклинические симптомы рака мозга вызвали требование к диагностическому сканированию, и именно они определяли появление рака мозга, а не облучение от Торотраста (хотя факт накопления указанного α -излучателя в тканях с последующей индукцией различных типов рака доказан достаточно давно).

Диагностическое облучение в стоматологии. В исследовании “случай–контроль” частоты менингиом после диагностического стоматологического облучения было обнаружено увеличение риска рака щитовидной железы. Но в таких опросных исследованиях сильно влияние смещения воспоминания (recall bias), которое, порой, полностью определяет регистрируемый канцерогенный эффект. Вероятной представляется обратная причинность: поскольку менингиомы способны вызывать боль в орофациальной области, пациенту с подобной болью могли проводить стоматологическую рентгеноскопию в течение курса его терапии.

Диагностическое облучение при компьютерной томографии (СТ). Начиная с 2012 г. в относительно небольшом числе (немногим более десятка) масштабных эпидемиологических исследований (10 стран) было заявлено о доказательствах значимого, нередко зависящего от дозы (=числа томографий) учащения раков и/или лейкозов после СТ в детском возрасте (только в двух работах эффекты не были выявлены). Однако, согласно ряду авторов (среди которых наиболее весома, возможно, публикация J.D. Voise Jr, одного из ведущих радиационных эпидемиологов США [139]), значительный, если не основной вклад может вносить обратная причинность или конфаундинг показания. Попытки устранить это временное смещение путем учета правдоподобия латентного периода не всегда оказывались удачными [139], и

потому почти во всех публикациях сами авторы указывают на возможность обратной причинности. Данный вопрос, понятно, имеет большую важность для здравоохранения и общественно-социальной сферы.

БИОЛОГИЧЕСКИЙ ГРАДИЕНТ ("BIOLOGICAL GRADIENT")

Неудовлетворительность термина

Это замысловатое название для зависимости "доза–эффект" присвоил критерию А.В. Hill в 1965 г. [3]. Хотя сам дозовый подход к подтверждению причинности связей был ясен уже давно. О зависимости "доза–эффект" сказано еще в Dorn H.F., 1953 ("...исследовать изменение степени ассоциации в связи с продолжительностью и интенсивностью курения") [2]. В 1964 г. A.J. Vollet прослеживал данный руководящий принцип вплоть до концепции клинической медицины от Г. Бурхава (Herman Boerhaave; 17–18 вв.) [2]. В том же 1964 г., в Сообщении Главного врача США о последствиях курения [20], среди пяти критериев не имелось зависимости от дозы, равно как и в соответствующем драфте материала от R.A. Stallones (1963) [2]. Но таковая была включена как бы подпунктом в критерии "Strength" и "Coherence", причем с обычным названием – "dose–effect" [20]⁴¹.

Термин "Биологический градиент", вероятно в связи с авторитетом Хилла, разошелся потом на полвека с лишним. Наш упоминавшийся выше анализ работ, в которых использовались критерии Хилла для доказательности (2013–2019), показал, что из 35 источников критерий зависимости от дозы применялся в 33, и в 26 из них (79%) он был назван строго "по Хиллу", порой – без добавления "доза–эффект".

Но для специальной литературы по теории причинности в эпидемиологии ситуация иная: в большинстве случаев термин все же заменен на "зависимость доза–эффект" и в этом роде. Однако и здесь множество исключений. Среди цитированных выше источников это работы (2005–2017) [19, 40, 72, 77, 104, 120], пособия по эпидемиологии (2007–2018) [30, 36, 67, 79, 82, 111, 135] и документы организаций США (2004–2005) [69, 97]. Есть и еще примеры.

Словом, Хилл умер (в 1991 г.), но дело его живет, порой даже в сомнительных мелочах. Иногда встречаются иные термины, к примеру "Exposure-varying association" [140], или, для модифицированных критериев Хилла в экоэпидемиологии, "Concordance of dose–response relationships" [140] (и мн. др. источники).

Иногда, что отмечалось выше, зависимость "доза–эффект" рассматривается как субкатегория критерия "Сила связи" [20, 33, 70, 71] (см. также в [7]).

Наличие дозовой зависимости усиливает причинный вывод

Исходно акцент в примерах был сделан на зависимость между числом выкуренных сигарет и частотой рака легкого; это рассматривается в большинстве источников на тему, с 1950-х годов [2] до 2018–2019 гг. [33, 36, 74].

Монотонные отношения между причинным фактором и его эффектом являются, по мнению M. Susser, своего рода стереотипом для осознания зависимости "доза–реакция"; они служат неким сигналом практически для всех исследователей [84]. Любой ответ, функционально связанный с дозой, является свидетельством возможной причинности [27, 36, 69, 72, 82, 85, 100, 110, 137]⁴².

Связь "доза–эффект" называют "самой фундаментальной и распространенной концепцией в токсикологии". Все вещества, даже вода, токсичны при достаточно высокой дозе. И наоборот, острые реакции связаны с пороговыми значениями; т.е. существует некоторая доза, ниже которой вероятность ответа индивидуума равна нулю [59]. Равным образом – и для фармакоэпидемиологии ("...является чрезвычайно важной и общепринятой концепцией в клинической фармакологии и аналогичным образом используется в эпидемиологии") [96]. Не исключение здесь и радиационная эпидемиология ("...выявление количественной связи между облучением и частотой заболевания; возможна зависимость от интенсивности или продолжительности радиационного воздействия" [131]). То же самое – для доказательной медицины: среди трех факторов, отвечающих за упоминавшийся ранее "Уровень достоверности доказательств", вторым называется "дозо-зависимый эффект" [46]; наличие такого эффекта повышает на один балл уровень достоверности клинической информации согласно международной системе оценки GRADE (Grading of Recommendations Assessment, Development, and Evaluation system) [141].

Для выявления дозовой зависимости на основе данных ряда публикаций возможно использование мета-анализа [26, 33, 142, 143].

В экоэпидемиологии рассматриваются дозовые ответы не только для конечного, но и для каждого промежуточного события в причинной цепочке ("key events") [140]. В молекулярной эпидемиологии критерий "Биологический градиент" предлагается расширить путем включения молекулярных связей "доза–ответ" [72].

Наличие зависимости "доза–эффект" не только обеспечивает следование биологической модели ассоциации, но и делает большинство не-причинных объяснений маловероятными. Это касается в основном случайностей, менее – смещений, но далеко не всегда – конфаундеров [36].

Конфаундеры могут имитировать зависимость “доза–эффект”

На такую возможность указано в ряде источников [27, 30, 33, 36, 69, 77, 111, 130, 144, 145] (и др.), но примеров приведено мало. Все они представлены ниже.

Помимо не раз рассматривавшейся связи между употреблением алкоголя и частотой рака легкого, где дозовую зависимость может имитировать сопутствующее курение [1, 27, 71, 146], называется также зависимость “доза–эффект” между курением и циррозом печени, которая имитируется уже употреблением алкоголя [130]. Сходная закономерность может быть и для иных заболеваний, обусловленных у курильщиков не только алкоголем, но и нездоровым питанием, недостаточным образованием и пр. [28].

Связь между порядком рождения ребенка у матери и частотой синдрома Дауна демонстрирует “биологический градиент”, однако — полностью объясняется конфаундером материнского возраста (см. в [1, 30]).

В публикациях К. Rothman и соавт. [77, 111] (и др.) обсуждается J-образная кривая зависимости “доза–эффект” между употреблением алкоголя и смертностью (т.е. не раз описанный факт большей смертности у полностью непьющих и переставших пить сравнительно с мало- и умеренно пьющими). Отмечается, что трудно судить, какие части этой кривой обусловлены некаузальными артефактами, вызванными конфаундерами или смещениями [77, 111]. В нашей работе [6] обсуждался этот вопрос в аспекте временного смещения. Специальное исследование [146] показало, что при корректировке эффекта на обратную причинность (учет в группе непьющих тех, кто ранее пил, но перестал; возможно, из-за состояния здоровья) и др. конфаундеры, “защитный эффект” для мало- и умеренно пьющих исчезает.

В Сообщении 1 [1] и публикации [6] нами приводились соответствующие примеры и из радиационной эпидемиологии:

- Ассоциация между дозой облучения и частотой рака легкого у работников ядерной индустрии США (имитация зависимости “доза–эффект” курением) [1].

- Ассоциация между дозой облучения и частотой курения у женщин, пострадавших от атомной бомбардировки в Хиросиме (возможность имитации зависимостей “доза–эффект” для последствий облучения курением) [1].

- Ассоциация между частотой раков и лейкозов у детей и дозой облучения, полученной их отцами — работниками ядерного центра в Селлафилде (имитация зависимости “доза–эффект” смешиванием популяции приезжими в ранний период (вирусные инфекции и т.д.) [1].

- Ассоциация между дозой облучения и частотой смертности от нерадиационно-обусловленных патологий (цирроза печени, рака ротовой полости и глотки, психозов и внешних причин) у работников ядерной индустрии Франции (имитация зависимости “доза–эффект” алкоголем) [1].

- Ассоциация между уровнем накопленной дозы облучения, с одной стороны, и курением, а также алкоголизмом, с другой стороны, у работников ядерной индустрии Японии. К тому же для более высоких дозовых групп частота прохождения рентгенодиагностики пищеварительного тракта и пр. была ниже [6, 147].

Функционально зависимость “доза–эффект” может быть какой угодно

Еще А.В. Hill в 1965 г. отмечал необязательность линейной или монотонной зависимости [3]. Этот критерий не основан на какой-то определенной форме дозовой кривой [36].

“Немонотонная связь только опровергает те причинные гипотезы, которые достаточно специфичны для прогнозирования монотонной кривой зависимости доза–реакция” [111]⁴³.

“Отсутствие правильной зависимости доза–эффект отвергает только гипотезы, предполагающие такую зависимость” [28].

Кривая “доза–эффект” отражает сложные биологические механизмы и, потому, может иметь любую форму [64, 145]. Критерий способен выполняться при пороговой [26, 30, 32, 33, 64, 71, 72, 74, 78, 79, 84, 110, 111, 135, 145, 148], сигмоидальной [110, 148] (S-образной [28, 84]), унимодальной [149], параболической [110, 148] (т.е. U-образной [72, 84]), гормезисной J-образной [28, 33, 71, 72, 77, 111, 150] и пр. функциях, но важно, чтобы ее можно было продемонстрировать (относительная интенсивность и длительность воздействия часто используются как суррогаты дозы [110]).

К примеру, зависимости “доза–эффект” между дозой радиации и частотой индуцированных облучением раков и лейкозов для средних (0.1–1.0 Гр [151]) и больших (свыше 1 Гр [151]) доз наилучшим образом описываются линейной и линейно-квадратичной функциями [13, 131]. Тканевые (детерминированные) эффекты излучения характеризуются порогами [50]. Для многих радиационных последствий показаны и гормезисные, J-образные кривые [152], хотя официальная позиция международных организаций состоит в том, что в аспекте эпидемиологии их однозначность не доказана [153] (дополнение к BEIR-VII [13])⁴⁴.

В то же время для области малых доз радиации с низкой ЛПЭ (до 0.1 Гр [13, 131, 151, 153]) в обсервационных, эпидемиологических исследованиях имеется масса неопределенностей, обусловлен-

ных обычно недостаточной статистической мощностью (невозможностью достижения в реале нужных величин выборок – sample size), влиянием множества нерадиационных факторов, конфаундеров и смещений. Поэтому международными и имеющими международный авторитет национальными организациями (НКДАР [131, 154], BEIR [13] и NCRP [155, 156]) признается для указанного диапазона равная вероятность любых форм зависимости от дозы: супралинейной (с гиперчувствительностью), линейной беспороговой, сублинейной или линейно-квадратичной, линейной с порогом и J-образной гормезисной (последняя не названа только в BEIR-VII [13]).

Неабсолютность критерия

Эпидемиологическое исследование может не обнаружить взаимосвязи между воздействием и реакцией по многим причинам (USEPA-2005 [97]). Поэтому отсутствие дозовой зависимости между фактором и реакцией не исключает причинно-следственную связь [32, 64, 77, 97, 137, 145] и не фальсифицирует ее [85].

Это все весомые ссылки: пособие по эпидемиологии от 2011 г. [32], авторитеты Rothman K.J., Greenland S., 2005 [77] и Susser M., 1986 [85], USEPA-2005 [97] и British Medical Association, 2004 [137]. Работы [64, 145] тоже серьезные: на стыке эпидемиологии и юриспруденции (критика правила Дауберта; см. выше) [64] и на стыке эпидемиологии и политики [145].

Однако в Hill A.V., 1965 [3] подобных категоричных заключений все же нет.

Следует иметь в виду также, что выявление зависимости “доза–эффект” может быть следствием самой возможности/невозможности исследовать такую зависимость (методической, ситуационной и пр.). Еще одна проблема имеет место, когда диапазон реально имеющихся (наблюдаемых) уровней экспозиции настолько мал, что нет способа зарегистрировать дозовый ответ вследствие, так сказать, “малой разрешимости” эксперимента [145].

Иногда наличие порога приравнивается к отсутствию дозовой зависимости (во всяком случае, так выглядит контекст) [78, 103, 145] (и др.), что верно только в случае последующего за порогом одинакового, “пологого” эффекта. На деле, к примеру, все тканевые эффекты облучения характеризуются и порогом, и, затем, четкими дозовыми зависимостями [50].

В некоторых источниках [64, 110, 111, 145] в контексте отсутствия дозовой зависимости рассматривается представленный выше пример с воздействием на беременных диэтилстилбестрола, который привел к аденокарциноме влагалища у их дочерей. Называется также и эффект индуцирования аденокарциномы печени винилхлоридом [64, 145]. На наш взгляд, учитывая яв-

ную недостаточность здесь выполнения критерия “Постоянство ассоциации” (исследования единичны), делать такие выводы нельзя.

В работе [157] указывается со ссылками, что дозовые ответы не были обнаружены, например, для легочных функций у астматиков после ингаляции кортикостероидов или при фармакотерапии психических заболеваний (последнее тоже несколько сомнительно). Сходным образом в [145] отмечается, что высыпания после воздействия вируса кори, равно как и титры соответствующих антител, могут не отвечать дозовой зависимости.

Более примеров на тему отсутствия дозовой зависимости при имеющейся причинности во многих десятках источников нами не найдено.

К сказанному надо заметить, что вывод о том, что для свидетельства о причинности никакая зависимость “доза–эффект” не нужна, в очередной раз дает широкое поле для разной сомнительной конъюнктуры, в том числе в радиационной сфере. Весомость приведенных выше соответствующих ссылок может усугублять ситуацию.

В этом плане вспоминаются массовые рассуждения 1990-х – первой половины 2000-х годов об отсутствии дозовой зависимости для радиационно-индуцированной нестабильности генома, что не отвечало действительности, а было обусловлено часто методическими артефактами или же неполнотой исследований (см. в [158]).

ЗАКЛЮЧЕНИЕ К ЧАСТИ 1

Представленное Сообщение 3 посвящено истории, сути и ограничениям наиболее известных эпидемиологических критериев причинности (“критерии Хилла”). Часть 1 охватила пять первых критериев из девяти: силу связи; постоянство ассоциации; специфичность; временную зависимость и биологический градиент (зависимость “доза–эффект”). Для каждого критерия были изложены теоретические и практические аспекты: история появления, терминология, философская и эпидемиологическая суть, применимость в различных дисциплинах (включая радиационную эпидемиологию) и ограничения. Только один из критериев может быть назван именно “критерием” – это временная зависимость (воздействие должно быть перед эффектом), поскольку фальсификация по нему (согласно подходам К. Поппера) устраняет причинность, так что никакие другие критерии уже не нужны. Остальные положения являются некими руководящими принципами (“viewpoints”, как назвал их А.В. Hill в 1965 г. [3]), выполнение которых повышает вероятность причинности ассоциации. Рассмотренные в части 1 Сообщение 3 критерии отвечают индуктивному, а не дедуктивному принципу научного познания, и потому, согласно положениям К. Поппера, не могут служить методологией истинной

доказательности. Однако реально, в практическом плане, все неэкспериментальные, наблюдательные дисциплины и подходы при поиске подтверждений причинности ассоциации опираются в первую очередь на указанные принципы, причем три из рассмотренных здесь пяти критериев (сила связи, постоянство ассоциации и зависимость “доза–эффект”) являются универсальными не только для эпидемиологии и иных наблюдательных дисциплин, но, стихийно, всегда принимаются во внимание и в лабораторном, и в ином эксперименте (а временная зависимость в контролируемом эксперименте выполняется априори).

Важным практическим аспектом изложенного материала является представление об общепринятых ординальных шкалах для градации силы ассоциации по разным параметрам. Таких шкал много, но, согласно нашему углубленному исследованию (выполненному, похоже, впервые не только для России, но и для иных стран) [7, 8], целесообразно придерживаться градаций, которые наиболее часто используются или отражены в весомых источниках. По показателю относительного риска (RR) основной следует считать шкалу Монсона (R. Monson; 1980–1990) [47], по отношению шансов (OR) – шкалу Розенталя (J.A. Rosenthal; 1996) [42]. Для предсказательной способности генетических маркеров градации OR можно найти в работе Рубанович А.В., Хромов-Борисов Н.Н., 2013; 2014 [56, 57]. А для коэффициента корреляции r Пирсона или Спирмена уместно использовать шкалу Хинкле и соавт. (Hinkle D.E., et al., 1979–2003) [76]. Последний показатель силы ассоциации мало информативен для эпидемиологии и молекулярной эпидемиологии, и наиболее часто используется в психосоциальных дисциплинах.

В части 2 Сообщения 3 намечено рассмотреть оставшиеся четыре критерия Хилла: биологическое правдоподобие, согласованность с текущими фактами и теоретическими знаниями, эксперимент и аналогию.

КОНФЛИКТ ИНТЕРЕСОВ И ВОЗМОЖНОСТЬ СУБЪЕКТИВНЫХ УКЛОНОВ

Конфликт интересов отсутствует. Представленное исследование, выполненное попутно в рамках более широкой бюджетной темы НИР ФМБА России, не поддерживалось никакими иными источниками финансирования. При выполнении работы не имелось временных рамок, официальных требований, ограничений, или же иных внешних объективных либо субъективных вмешивающихся факторов.

ПРИМЕЧАНИЯ

1. Не только в описательных с отсутствием адекватного контроля, но даже в экспериментальных исследованиях с какими угодно контролями, с философской позиции нельзя получить строгие доказательства эффекта от воздействия или следствия от причины. Время не течет вспять, и никто не может наблюдать одно и то же событие, одного и того же индивидуума, или же одну и ту же популяцию (либо группу лабораторных животных) дважды – с конкретным условием (“причиной”) и – без него, причем одновременно. Этот недостижимый сценарий называется “контрафактический идеал” и “контрафактический контраст” [10]. Все, что рассматривается как экспериментальный подход, по сути основано на суррогате контрафактического [10] субъекта/объекта, который и называют “контролем” [10, 19]. Контрафактическая концепция рассмотрена в нашем обзоре [10].

2. На заре исследования СПИДа его причины связали с употреблением гомосексуалистами амилнитритов (“попперсов”; вдыхание их увеличивает, скажем так, временный темперамент). Относительный риск (RR) составил 12.3 [29]. Такая величина ассоциации должна быть названа “очень сильной” по каким угодно градациям [7] (см. также ниже). Сходным образом, стандартный пример конфаундера – связь между употреблением алкоголя и раком легкого (курение), тоже может быть сильной, равно как ассоциация между курением и циррозом печени [1]. Значительная величина связи показана между порядком рождения ребенка и частотой синдрома Дауна, что, понятно, обусловлено просто возрастом матери [30].

3. В качестве примера относительно умеренной, но важной для общества ассоциации называется связь между активным курением и сердечно-сосудистыми патологиями (частота смертности увеличена для разных заболеваний в 1.5–2.6 раза [20]) [14, 20, 28, 30, 31] (и многие другие), а также между пассивным курением и ишемической болезнью сердца [33]. Прирост во втором случае от фоновой частоты вовсе невелик (RR = 1.2–1.3), но фоновое значение – высокое [33]. Другим примером “социально-значимой” слабой ассоциации является связь между пассивным курением и раком легкого. В течение десятков лет, путем значительных усилий и множества работ, реализовавшихся в ряде мета-анализов, удалось в конце концов показать, что у жен курильщиков RR для рака легкого составляет 1.2–1.25 (к 2002–2006 гг.) [34–36]. Отметим здесь, что фоновая частота рака легкого у некурящих несравнима с таковой для ишемической болезни сердца. Поэтому один и тот же RR порядка 1.2 (по количественным понятиям риска в эпидемиологии подобная связь вообще недоказуема [7]; см. также ниже) для двух названных патологий означает совсем разный аб-

солютный прирост, во втором случае ничтожный. В [34] указано, что связь между пассивным курением и раком легкого — один из немногих примеров, когда сообществом эпидемиологов была принята столь слабая ассоциация за реальную. Исследования пассивного курения вообще чрезвычайно трудно спроектировать и интерпретировать [37]. Так, в [38] была показана слабая связь между пассивным курением родителей и плохим состоянием здоровья у детей с муковисцидозом. Здесь, однако, возможно влияние неучтенных социально-экономических факторов: насколько родители заботятся о таких своих детях, в том числе решили ли они перестать курить для них [22].

4. В 1960-х годах статистик в области психологии и социологии Jacob Cohen (1923–1998; США) для величины эффекта (“effect size”) применительно к связи между переменными, предложил количественные градации значений для ряда параметров: d, r, q, h, w, fi, f^2 . Эти параметры отражают, соответственно: t -тест для средних, корреляцию Пирсона, различие между коэффициентами корреляции, различия между пропорциями, χ^2 -тест на распределение, анализ на дисперсию и ковариацию, а также на множественную регрессию [41]. Позднее значения для диапазонов величины эффекта по тому или иному параметру предлагались еще многими авторами [7, 8].

5. J. Cornfield на основе связи между курением и раком легкого путем расчетов вывел некое неравенство между величиной эффекта исследуемого воздействия и эффекта возможного конфаундера (“Cornfield’s Inequality”), позволяющее делать вероятный вывод о наличии истинной причинности. Была оценена минимальная величина эффекта конфаундера, который мог бы аннулировать или полностью изменить интересующую зависимость [7]. Но откуда взялось значение именно $RR < 3$ для границы “слабой связи”? Объяснений нам найти не удалось ни в работах 1950-х годов, ни позднее (только констатацию). Можно предположить следующее. Ранее мы рассматривали [2] “диссидентские” попытки одного из наиболее известных статистиков, Р.А. Фишера (R.A. Fisher; F -критерий; уровень значимости $p = 0.05$ и др.) опровергнуть причинность связи между курением и раком легкого. Одной из гипотез была генетическая обусловленность рака легкого (т. е. обратная причинность; см. в [1, 6]). Опыты Р.А. Фишера на близнецах продемонстрировали вклад генетического фактора, равный 3.0. Однако J. Cornfield указал Р. Фишеру, что RR для рака легкого у курильщиков составляет от 9 и более (до 20–60 в зависимости от числа выкуренных сигарет [2, 3, 14]). Таким образом, эффект не мог быть объяснен конфаундером [7]. Возможно, с тех пор величина $RR = 3.0$ и осталась граничной для поправки на иные факторы, хотя это только наше предположение.

6. “Мы ищем RR от трех или более”; “Если RR не будет равен, по крайней мере, трем или четырем, то забудьте о нем”; “Если это RR , равный 1.5, но он показан только в одном исследовании, даже очень хорошо, то вы... скажете: “Может быть”” (все — 1995 г.); “В Сообщениях Главного врача США о последствиях курения от 1964 г. и 1982 г. не было выявлено никаких утверждений о причинных зависимостях для $OR < 3.0$ ” (источник 2006 г.); “ OR ниже 3.0 редко указывает на клинически значимое открытие” (2012). Оригинальные цитаты и ссылки см. в [7]. Сходные построения есть в пособии по доказательной медицине ее основоположника D. Sackett и соавт. (1997) [44], где указывается, что для дизайна “случай–контроль” граница “впечатляющего эффекта” может быть при $RR = 4$, а для когортных исследований — при $RR = 3$. Это же повторяется в пособии по доказательной медицине от 2019 г. [45].

7. Второе издание Monson R.R., 1990 [47] частично представлено в виде Google-book. Материал со шкалой Монсона в оригиналах 1980 и 1990 г. был нам недоступен, но шкала реконструировалась по семи независимым источникам 1985–2015 гг. [7].

8. По словам С.Ж. Ferguson [43], в результате психологам присуще “чувство незащищенности” от восприятия психологии как “облегченной науки”, в то время как медицинские исследования воспринимаются ими как “серьезная наука”. Иной раз, вследствие трудностей установления причинности, психологию с трудом относят к реальной науке [54]. Это объясняется тем, что в психосоциологии исследуются почти только корреляции: “В психиатрической эпидемиологии наша научная база хрупка. Чтобы обнаружить причинные связи, мы вынуждены полагаться на качество и количество ассоциаций, которые и обсуждаются” [55] (оригинальные цитаты см. в [8]). А примеры того, к чему приводит опора только на ассоциации, были приведены нами ранее [1, 6, 7]. Среди них — значимая корреляция между потреблением мороженого и смертностью от утопления [1], между числом фильмов по годам, в которых сыграл Н. Кейдж, и числом утонувших в эти годы в бассейнах США [6], или строгая обратная зависимость между величиной валового внутреннего продукта (ВВП) в стране и линейным размером пениса у ее мужского населения (выборка по 76 странам всех континентов; не пародия и не от 1 апреля, а рядовая среди прочих публикация (с пометкой “дискуссионная”) с Helsinki Center of Economic Research. Автор связал с уровнем тестостерона склонность к рискам, в том числе экономическим, а про расовый фактор забыл; подробности см. в [7]).

9. “Today, statistical significance — not the magnitude of association — is the accepted benchmark for judging the strength of an observed association, and thus its potential causality” [72].

10. “Мера силы ассоциации – это относительный риск, а не статистическая значимость”. (“The measure of strength of an association is the relative risk and not statistical significance.”) [67]. “Мы утверждаем, что научная информативность причинной гипотезы не связана с величиной ассоциации” (“We contend that the scientific informativeness of a causal hypothesis is not related to the magnitude of the association”) [73].

11. Квадрат коэффициента корреляции r Пирсона (R^2), так называемый “коэффициент детерминации”, отражает долю дисперсии признаков (в %), обусловленную линейной регрессией, т.е. это процент от варьирования одной переменной, связанный с варьированием другой согласно линейной модели [28, 41]. Расчет показывает, что даже при $r = 0.5$ (по стихийным понятиям это немалая корреляция) доля связанной дисперсии составляет всего четверть от максимальной (0.25). Более половины вариаций (>50%), т.е. когда коррелятивная связь начинает превалировать над ее отсутствием (по аналогии с правилом Дауберта, см. выше), можно ожидать только при $r > 0.7$.

12. Это коснулось и нас в Сообщении 2. Слово “consistent” нашлось в публикациях Wynder E.L., 1956; 1961 [2] в пункте правил установления причинности патологий. Мы неправомерно отнесли пункт к критерию “Consistency”, хотя смысл был иной – согласованность/совпадение в популяции распределения патологии и ее возможной причины. В результате работам Wynder E.L., 1956; 1961 в табл. 2 из [2] нами было приписано предложение в том числе критерия “Consistency”. Хотя в 1959 г., как было видно, А.М. Lilienfeld [2] все-таки связал подобную “согласованность” с постоянством ассоциации.

13. Во всех известных нам крупных отечественных пособиях по эпидемиологии и доказательной медицине (см. в [6–10]) среди критериев Хилла “Coherence” отсутствует. Но в двух более кратких руководствах авторы все же приводят данный пункт, возможно потому, что ничего не меняли, как в первых случаях, от себя. Назовем здесь такие источники: пособия Первого МГМУ им. И.М. Сеченова (экологические заболевания; Семеновых Г.К. и др., 2011 [80]) и КГМУ (эпидемиология; Хасанова Г.Р. и др., 2017 [81]).

14. “...who believe that we all use induction every day to negotiate through every aspect of our environment”) [92].

15. “Inductive methods are part of the framework of our life, of human nature”) [93].

16. “Justifying induction would be pointless because inductive arguments serve to define what rational arguments are all about”) [93].

17. “If we truly limit ourselves to deductive reasoning, we will be forced to throw away most of our beliefs about the world and to give up all hope of knowledge

of nature, not to mention any hope for improving peoples' health”) [94].

18. “Обыкновенно признавали, что для осуждения человека, пользующегося хорошей репутацией, необходимы два свидетеля, хотя некоторые инквизиторы требовали большего числа”. (“Two witnesses were usually assumed to be necessary for the condemnation of a man of good repute, though some authorities demanded more.”) [95].

19. “...a single study is seldom, if ever, conclusive” [84].

20. “...one should never believe a finding reported only once: there may have been an error committed in the study which is not apparent to either the investigator or the reader” [96].

21. “The opportunities for bias are so many that one study is rarely conclusive” [22].

22. “...there will be occasions when repetition is absent or impossible and yet we should not hesitate to draw conclusions” [3].

23. “Not every hypothesis affords the opportunity for the replication of studies” [84].

24. “In my view, consistency is the most powerful verification available” [85].

25. “The reproducibility of findings constitutes one of the strongest arguments for causality” [97].

26. Здесь и далее в перечислении приведены ссылки на узкую тему: устранение путем повторов исследований того или иного конкретного типа искажающих факторов.

27. Как указано в [99], в разных странах такие факторы, как самооценка стресса, диета или величина новорожденного могут быть по-разному связаны с социоэкономическими условиями и социоэкономическими составляющими патологий. Поэтому выявление ассоциации в разных популяциях дает некоторую защиту от эффекта конфаундеров.

28. Видно, что ни в одном источнике нет полного перечисления всех пунктов, предложенных с 1960-х годов “коллективным разумом”, хотя к этому приближается оксфордский словарь по эпидемиологии [15] и работа Susser M., 1986 [84].

29. “There are no fixed guidelines on how many studies need to be done, how diverse they need to be, and how relevant they are to the question at hand to guarantee the benefits of consistency. These are matters of scientific judgment; explicit criteria cannot be offered” [36]. К.Ж. Goodman с сотр. (J.M. Samet [36] или С.В. Phillips [101]), являются, похоже, “штатными” авторами статей и разделов по причинности эффектов в огромных западных пособиях и энциклопедиях, от эпидемиологии до психологии; много источников (2004–2018 гг., например, [36, 101]).

30. Если представить себе исследования связи между алкоголизмом у мужчин и увеличением частоты рака легкого, без поправки на курение, то,

вероятно, любые повторы по всем странам даже за 100 лет, равно как и соответствующие мета-анализы, не смогут устранить эффект конфаундера. То же самое можно сказать, по-видимому, и про учашение раков и лейкозов после компьютерной томографии в детском возрасте; устранить в этих случаях вклад обратной причинности трудно (см. в [6] и ниже).

31. World Health Organization/International Programme on Chemical Safety. Обеспечивает официальную структуру для оценки данных о направлениях (pathways) причинно-следственных ключевых событий, ведущих к неблагоприятным последствиям для здоровья [105].

32. О периодах индукции эффекта, латентном и инкубационном, подробнее см. в [6].

33. Хотя даже для некоторых инфекционных заболеваний нет возможности выявить специфический вирулент. В Сообщении 2 [2] уже указывалось, что, согласно [18], никакой возбудитель не может быть идентифицирован у 25–50% больных острым респираторным синдромом, у порядка 75% пациентов с острыми лихорадочными синдромами с участием ЦНС или в большинстве случаев общего острого гастроэнтерита.

34. “When it is met, however, it provides extremely strong support for a conclusion that an association is causal” [96].

35. Согласно рассуждениям К. Поппера, если нам встречаются только белые лебеди, то, сколько бы их ни было, нельзя делать индуктивного вывода, что все лебеди – белые. Вдруг где-то существуют небелые лебеди (и они действительно есть – черные). М. Susser развил это рассуждение на возможность нахождения небелого полярного медведя [85].

36. “The report [1964] notes the extremely high relative risk for lung cancer in smokers and the high attributable risk, and concludes that the association between smoking and lung cancer has “a high degree of specificity” [69].

37. “Если другие причины смерти повышаются у курящих на 10, 20 или даже на 50%, то рак легкого повышается на 900–1000%, и у нас есть специфичность – специфичность в величине ассоциации”. (“If other causes of death are raised 10, 20 or even 50% in smokers whereas cancer of the lung is raised 900–1000% we have specificity – a specificity in the magnitude of the association”.) [3].

38. “True associations are specific, and in most circumstances it is not plausible that a drug is associated with an increase in developing multiple cancers” [129].

39. Мы все же поправим [103], что аномалии могут быть разными по тяжести и, поэтому, по произведенному когда-то на опрашиваемого впечатлению, что способно отразиться на воспоминаниях.

40. “I believe that multi-causation is generally more likely than single causation though possibly if we knew all the answers we might get back to a single factor” [3].

41. “Important to the strength as well as to the coherence of the association is the dose–effect phenomenon” [20].

42. Действительно, для проводящего эксперименты на животных в лаборатории, что предусматривает контрольные группы, критерии причинности и доказательности в описательных дисциплинах (и в жизни) обычно не ближе китайской грамоты (кроме, понятно, Китая). Но вспоминается, что два критерия обязательны и тут, во многом по стихийным соображениям. Один из них как раз – зависимость “доза–эффект”. Второй, “контрафактический эксперимент” [1, 10], заключающийся в попытке направленно подавить воздействие фактора (элиминирование, ингибирование, блокирование, нокаутирование и др.), а затем проследить за эффектом, все же вторичен сравнительно с первым.

43. “A nonmonotonic relation only refutes those causal hypotheses specific enough to predict a monotonic dose–response curve” [111].

44. “Обсуждение радиационного гормезиса продолжается... Комитет BEIR VII заключил, что если уровни воздействия или кривые [дозового] ответа на воздействие различаются для особей в популяции, или же ситуация непонятна, влияние гормезиса на оценку риска малых доз требует дополнительного исследования и остается неопределенным. Несмотря на рост числа публикаций и информированности в этой области [со времени издания BEIR-VII], конкретные доказательства, которые позволили бы изменить названное заключение BEIR-VII, не представляются существенными” (2010). (“The discussion on radiation hormesis continues... The BEIR VII committee concluded that while exposure levels or exposure response functions differ between individuals in a population or are uncertain, the impact of hormesis on low-dose risk requires additional research and remains uncertain. Despite the increasing number of publications and awareness in this area, there does not appear to be insufficient concrete evidence to deviate from BEIR VII’s conclusion at this time”.) [153].

СПИСОК ЛИТЕРАТУРЫ

1. Котеров А.Н. Критерии причинности в медико-биологических дисциплинах: история, сущность и радиационный аспект. Сообщение 1. Постановка проблемы, понятие о причинах и причинности, ложные ассоциации // Радиационная биология. Радиоэкология. 2019. Т. 59. № 1. С. 1–32. [Kotero A.N. Causal criteria in medical and biological disciplines: history, essence and radiation aspect. Report 1. Problem statement, conception of causes and causation, false associations // Radiat. Biol. Radioecol. (Radiation Biology and Ecology) 2019. Vol. 59. No. 1. P. 1–32.]

- tion biology. *Radioecology*, Moscow). 2019. V. 59. № 1. P. 1–32. (In Russian. Engl. abstr.)
<https://doi.org/10.1134/S0869803119010065>
2. *Котеров А.Н.* Критерии причинности в медико-биологических дисциплинах: история, сущность и радиационный аспект. Сообщение 2. Постулаты Генле–Коха и критерии причинности неинфекционных патологий до Хилла // *Радиационная биология. Радиоэкология*. 2019. Т. 59. № 4. С. 341–375. [*Kotero A.N.* Causal criteria in medical and biological disciplines: history, essence and radiation aspect. Report 2. Henle–Koch postulates and criteria for causality of non-communicable pathologies before Hill // *Radiats. Biol. Radioecol.* (Radiation biology. Radioecology, Moscow). 2019. V. 59. № 4. P. 341–375. (In Russian. Engl. abstr.)]
<https://doi.org/10.1134/S0869803119040052>
 3. *Hill A.B.* The environment and disease: association or causation? // *Proc. R. Soc. Med.* 1965. V. 58. № 5. P. 295–300.
<https://doi.org/10.1177/0141076814562718>
 4. *Good I.J.* Weight of evidence, corroboration, explanatory power, information, and the utility of experiments // *J. Royal Stat. Soc. Series B: Methodological*. 1960. V. 22. № 2. P. 319–331.
<https://doi.org/10.1111/j.2517-6161.1960.tb00378.x>
 5. *Martin P., Bladier C., Meek B. et al.* Weight of evidence for hazard identification: a critical review of the literature // *Environ. Health Perspect.* 2018. V. 126. № 7. Art. 076001.
<https://doi.org/10.1289/EHP3067>
 6. *Котеров А.Н., Ушенкова Л.Н., Бирюков А.П.* Критерий Хилла “Временная зависимость”. Обратная причинность и ее радиационный аспект // *Радиационная биология. Радиоэкология*. 2020. Т. 60. № 2. С. 115–152. [*Kotero A.N., Ushenkova L.N., Biryukov A.P.* Hill’s criteria “Temporality”. Reverse causation and its radiation aspect // *Radiats. Biol. Radioecol.* (Radiation biology. Radioecology, Moscow). 2020. V. 60. № 2. P. 115–152. (In Russian. Engl. abstr.)]
<https://doi.org/10.31857/S086980312002006X>
 7. *Котеров А.Н., Ушенкова Л.Н., Зубенкова Э.С. и др.* Сила связи. Сообщение 1. Градации относительного риска // *Мед. радиология и радиац. безопасность*. 2019. Т. 64. № 4. С. 5–17. [*Kotero A.N., Ushenkova L.N., Zubenkova E.S. et al.* Strength of association. Report 1. Graduation of relative risk // *Medit. Radiolog. Radiat. Bezopasnost* (Medical Radiology and Radiation Safety; Moscow). 2019. V. 64. № 4. P. 5–17. (In Russian. Engl. abstr.)].
https://doi.org/10.12737/article_5d1adb25725023.14868717
 8. *Котеров А.Н., Ушенкова Л.Н., Молодцова и др.* Сила связи. Сообщение 2. Градации величины корреляции // *Мед. радиология и радиац. безопасность*. 2019. Т. 64. № 6. С. 12–24. [*Kotero A.N., Ushenkova L.N., Molodtsova D.V. et al.* Strength of association. Report 2. Graduation of correlation size // *Medit. Radiolog. Radiat. Bezopasnost* (Medical Radiology and Radiation Safety; Moscow). 2019. V. 64. № 6. P. 12–24. (In Russian. Engl. abstr.)]
<https://doi.org/10.12737/1024-6177-2019-64-6-12-24>
 9. *Котеров А.Н., Ушенкова Л.Н., Бирюков А.П.* Критерий Хилла “Биологическое правдоподобие”. Интеграция данных из различных дисциплин в эпидемиологии и радиационной эпидемиологии // *Радиационная биология. Радиоэкология*. 2020. Т. 60. № 5. С. 453–480. [*Kotero A.N., Ushenkova L.N., Biryukov A.P.* Hill’s criteria “Biological plausibility”. The data integration from different disciplines in Epidemiology and Radiation Epidemiology // *Radiats. Biol. Radioecol.* (Radiation biology. Radioecology, Moscow). 2020. V. 60. № 5. P. 453–480. (In Russian. Engl. abstr.)]
<https://doi.org/10.31857/S0869803120050069>
 10. *Котеров А.Н., Ушенкова Л.Н., Бирюков А.П.* Критерий Хилла “Эксперимент”. Контрафактический подход в дисциплинах нерадиационного и радиационного профиля // *Радиационная биология. Радиоэкология*. 2020. Т. 60. № 6. С. 565–594. [*Kotero A.N., Ushenkova L.N., Biryukov A.P.* Hill’s criteria “Experiment”. The counterfactual approach in non-radiation and radiation sciences // *Radiats. Biol. Radioecol.* (Radiation biology. Radioecology, Moscow). 2020. V. 60. № 6. P. 565–594. (In Russian. Engl. abstr.)]
<https://doi.org/10.31857/S0869803120060193>
 11. *Weed D.L., Hursting S.D.* Biologic plausibility in causal inference: current method and practice // *Am. J. Epidemiol.* 1998. V. 147. № 5. P. 415–425.
<https://doi.org/10.1093/oxfordjournals.aje.a009466>
 12. *Weed D.L.* Analogy in causal inference: rethinking Austin Bradford Hill’s neglected consideration // *Ann. Epidemiol.* 2018. V. 28. № 5. P. 343–346.
<https://doi.org/10.1016/j.annepidem.2018.03.004>
 13. BEIR VII Report 2006. Phase 2. Health Risks from Exposure to Low Levels of Ionizing Radiation. Committee to Assess Health Risks from Exposure to Low Levels of Ionizing Radiation, National Research Council. <http://www.nap.edu/catalog/11340.html> (address data 11.10.2020).
 14. *Doll R.* Weak associations in epidemiology: importance, detection, and interpretation // *J. Epidemiol.* 1996. V. 6. № 4. Suppl. P. S11–S20.
https://doi.org/10.2188/jea.6.4sup_11
 15. *A Dictionary of Epidemiology* / Ed. J.M. Last. Oxford: Oxford University Press, 2001. [Эпидемиологический словарь / Под ред. Дж.М. Ласта для Международной эпидемиологической ассоциации: Пер. с англ. под ред. В.В. Власова (отв. ред.) и др. М.: Открытый институт здоровья в рамках проекта “Глобус”, 2009. 316 с.]
 16. *Susser M.* Causal Thinking in the Health Sciences: Concepts and Strategies of Epidemiology. New York: Oxford Univ. Press, 1973. 181 p.
 17. *Susser M.* What is a cause and how do we know one? A grammar for pragmatic epidemiology // *Am. J. Epidemiol.* 1991. V. 133. № 7. P. 635–648.
<https://doi.org/10.1093/oxfordjournals.aje.a115939>
 18. *Evans A.S.* Causation and disease: The Henle-Koch postulates revisited // *Yale J. Biol. Med.* 1976. V. 49. № 2. P. 175–195.
 19. *Bae S., Kim H.C., Ye B. et al.* Causal inference in environmental epidemiology // *Environ. Health Toxicol.*

2017. V. 32. Art. e2017015.
<https://doi.org/10.5620/eh.t.e2017015>
20. United States Department of Health, Education and Welfare (USDHEW). Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service Publication No. 1103. Washington DC: U.S. Department of Health, Education and Welfare. 1964. 387 p. <https://biotech.law.lsu.edu/cases/tobacco/nbbmq.pdf> (Address data 11.10.2020).
 21. *Weed D.L., Gorelic L.S.* The practice of causal inference in cancer epidemiology // *Cancer Epidemiol. Biomarkers Prev.* 1996. V. 5. № 4. P. 303–311.
 22. *Angell M.* The interpretation of epidemiologic studies // *N. Engl. J. Med.* 1990. V. 323. № 12. P. 823–825. <https://doi.org/10.1056/NEJM199009203231209>
 23. *Glasziou P., Chalmers I., Rawlins M., McCulloch P.* When are randomised trials unnecessary? Picking signal from noise // *Brit. Med. J.* 2007. V. 334. № 7589. P. 349–351. <https://doi.org/10.1136/bmj.39070.527986.68>
 24. *Doll R., Peto R.* Randomised controlled trials and retrospective controls // *Brit. Med. J.* 1980. V. 280. № 6206. P. 44. <https://doi.org/10.1136/bmj.280.6206.44>
 25. *Rawlins M.* De Testimonio: on the evidence for decisions about the use of therapeutic interventions // *Clin. Med. (Lond).* 2008. V. 8. № 6. P. 579–588. <https://doi.org/10.7861/clinmedicine.8-6-579>
 26. Handbook of Epidemiology / Eds W. Ahrens, I. Pigeot. 2nd Ed. New York, Heidelberg, Dordrecht, London: Springer, 2014. 2498 p.
 27. *Epidemiology: Principles and Practical Guidelines* / Eds J. Van den Broeck, J.R. Brestoff. Dordrecht: Springer, 2013. 621 p.
 28. *Власов В.В.* Эпидемиология: учебное пособие. 2-е изд., испр. М.: ГЭОТАР-Медиа, 2006. 464 с. [*Vlasov V.V.* Epidemiology. 2nd ed., revis. Moscow: GEOTAR-Media, 2006. 464 p. (In Russian)]
 29. *Phillips A.N., Davey Smith G.* Confounding in epidemiological studies // *Brit. Med. J.* 1993. V. 306. № 870. P. 142. <https://doi.org/10.1136/bmj.306.6870.142-b>
 30. *Rothman K.J.* Epidemiology. An Introduction. 2nd ed. Oxford etc.: Oxford University Press Inc., 2012. 268 p.
 31. *Rothman K.J., Poole C.* A strengthening programme for weak associations // *Int. J. Epidemiol.* 1988. V. 17. № 4. P. 955–959. <https://doi.org/10.1093/ije/17.4.955>
 32. *Webb P., Bain C.* Essential Epidemiology. An Introduction for Students and Health Professionals. 2nd ed. Cambridge etc.: Cambridge University Press, 2011. 445 p.
 33. *Szklo M., Nieto F.J.* Epidemiology. Beyond the Basics. 4th ed. Burlington: Jones & Bartlett Learning, 2019. 577 p.
 34. *Boffetta P.* Causation in the presence of weak associations // *Crit. Rev. Food Sci. Nutrit.* 2010. V. 50. № S1. P. 13–16. <https://doi.org/10.1080/10408398.2010.526842>
 35. *Boffetta P.* Involuntary smoking and lung cancer // *Scand. J. Work Environ. Health.* 2002. V. 28. Suppl. 2. P. 30–40.
 36. *Goodman S.N., Samet J.M.* Cause and Cancer Epidemiology // Schottenfeld and Fraumeni Cancer Epidemiology and Prevention. 4th ed. / Eds M.J. Thun et al. New York: Oxford University Press. Printed by Sheridan Books, Inc., USA, 2018. P. 97–104.
 37. *Rylander R.* Environmental tobacco smoke and lung cancer // *N. Engl. J. Med.* 1990. V. 323. № 12. P. 834. <https://doi.org/10.1056/NEJM199009203231217>
 38. *Rubin B.K.* Exposure of children with cystic fibrosis to environmental tobacco smoke // *N. Engl. J. Med.* 1990. V. 323. № 12. P. 782–788. <https://doi.org/10.1056/NEJM199009203231203>
 39. *Brown S.M.* The use of epidemiologic data in the assessment of cancer risk // *J. Environ. Pathol. Toxicol.* 1980. V. 4. № 2, 3. P. 573–580.
 40. *Lower G.M., Kanarek M.S.* Conceptual/operational criteria of causality: relevance to systematic epidemiologic theory // *Med. Hypothes.* 1983. V. 11. P. 217–244. [https://doi.org/10.1016/0306-9877\(83\)90064-6](https://doi.org/10.1016/0306-9877(83)90064-6)
 41. *Cohen J.* Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, Mahwah, NJ: Lawrence Erlbaum Associates, 1988. 567 p.
 42. *Rosenthal J.A.* Qualitative descriptors of strength of association and effect size // *J. Soc. Serv. Res.* 1996. V. 21. № 4. P. 37–59. https://doi.org/10.1300/J079v21n04_02
 43. *Ferguson C.J.* Is psychology research really as good as medical research? Effect size comparisons between psychology and medicine // *Rev. Gen. Psychol.* 2009. V. 13. № 2. P. 130–136. <https://doi.org/10.1037/a0015103>
 44. *Sackett D.L., Richardson W.S., Rosenberg W., Haynes R.B.* Evidence-Based Medicine: How to Practice and Teach EBM. New York: Churchill Livingstone, 1997. 250 p.
 45. *Straus S.E., Glasziou P., Richardson W.S. et al.* Evidence-Based Medicine: How to Practice and Teach EBM. 5th ed. Edinburgh etc.: Elsevier, 2019. 406 p.
 46. *Андреева Н.С., Реброва О.Ю., Зорин Н.А. и др.* Системы оценки достоверности научных доказательств и убедительности рекомендаций: сравнительная характеристика и перспективы унификации // *Мед. технологии. Оценка и выбор.* 2012. № 4. С. 10–24. [*Andreeva N.S., Rebrova O.Y., Zorin N.A. et al.* Systems for assessing the reliability of scientific evidence and the soundness of guidelines: comparison and prospects for unification // *Medical Technologies. Assessment and Choice (Moscow).* 2012. № 4. P. 10–24. (In Russian. Engl. abstr.)]
 47. *Monson R.R.* Occupational Epidemiology. 2nd ed. Florida: Boca Raton, CRC Press Inc., 1990. 312 p.
 48. *Singer R.* Introduction to epidemiology and causal inference. Presentation // Executive Veterinary Program. University of Illinois, 2014. 8 p. <http://vetmed.illinois.edu/wp-content/uploads/2015/10/6.-Module-2.1.pdf> (address data 25.01.2019); <https://vetmed.umn.edu/bio/vmed-grad-faculty/randall-singer> (address data 15.10.2020).

49. *Little M.P., Azizova T.V., Bazyka D. et al.* Systematic review and meta-analysis of circulatory disease from exposure to low-level ionizing radiation and estimates of potential population mortality risks // *Environ. Health Perspect.* 2012. Vol. 120. № 11. P. 1503–1511. <https://doi.org/10.1289/ehp.1204982>
50. ICRP Publication 118. ICRP Statement on tissue reactions and early and late effects of radiation in normal tissues and organs – threshold doses for tissue reactions in a radiation protection context. *Annals of the ICRP / Ed. C.H. Clement.* Amsterdam–New York: Elsevier, 2012. 325 p.
51. *Lagarde F., Cullings H.M., Shimizu Y., Cologne J.B.* Tiny excess relative risks hard to pin down (rapid response to Cardis et al 2005) // *Brit. Med. J.* 2005. V. 331. P. 77–81. <https://doi.org/10.1136/bmj.38499.599861.E0> <https://www.bmj.com/content/331/7508/77/rapid-responses> (address data 11.10.2020)
52. *Cardis E., Vrijheid M., Blettner M. et al.* Risk of cancer after low doses of ionising radiation: retrospective cohort study in 15 countries // *Brit. Med. J.* 2005. V. 331. № 7508. P. 77–82. <https://doi.org/10.1136/bmj.38499.599861.E0>
53. *Lagarde F.* Methodology issues in epidemiological assessment of health effects of low-dose ionising radiation // *Radiat. Prot. Dosim.* 2003. V. 104. № 4. P. 297–314. <https://doi.org/10.1093/oxfordjournals.rpd.a006193>
54. *Ferguson C.J.* “Everybody knows psychology is not a real science”: Public perceptions of psychology and how we can improve our relationship with policymakers, the scientific community, and the general public // *Am. Psychologist.* 2015. V. 70. № 6. P. 527–542. <https://doi.org/10.1037/a0039405>
55. *Schwab J.J., Schwab M.E.* Sociocultural Roots of Mental Illness. An Epidemiologic Survey. New York: Springer US, 1978. 338 p.
56. *Рубанович А.В., Хромов-Борисов Н.Н.* Теоретический анализ показателей предсказательной эффективности бинарных генетических тестов // *Экол. генетика.* 2013. Т. 11. № 1. С. 77–90. [*Rubanovich A.V., Khromov-Borisov N.N.* Theoretical analysis of the predictability indices of the binary genetic tests // *Ecological Genetics.* 2013. V. 11. № 1. P. 77–90. (In Russian. Engl. abstr.)]. <https://doi.org/10.17816/ecogen11177-90>
57. *Rubanovich A.V., Khromov-Borisov N.N.* Theoretical analysis of the predictability indices of the binary genetic tests // *Russian Journal of Genetics: Appl. Res.* 2014. V. 4. № 2. P. 146–158. <https://doi.org/10.1134/S2079059714020087>
58. *Little J.B.* Radiation carcinogenesis // *Carcinogenesis.* 2000. V. 21. № 3. P. 397–404. <https://doi.org/10.1093/carcin/21.3.397>
59. *Ulsh B.A.* Are risks from medical imaging still too small to be observed or nonexistent? // *Dose Response.* 2015. V. 13. № 1. Article pii: dose-response.14-030.Ulsh. <https://doi.org/10.2203/dose-response.14-030.Ulsh>
60. *Royal H.D.* Good News About Radiation Effects??? Editorial // *Am. J. Med.* 2016. V. 129. № 6. P. 549–550. <https://doi.org/10.1016/j.amjmed.2015.12.002>
61. IARC 2012. Radiation. A review of human carcinogens. V. 100 D. IARC monographs on the evaluation of carcinogenic risks to humans. Lyon, France, 2012. 341 p.
62. *Hunter R.J., Jr., Shannon J.H., Amoroso H.J.* How to manage issues relating to the use of trial experts: standards for the introduction of expert testimony through judicial “Gate-Keeping” and scientific verification // *J. Management and Strategy.* 2018. V. 9. № 1. 11 p. <https://doi.org/10.5430/jms.v9n1p1>
63. *Cole P.* Causality in epidemiology, health policy and law // *Environ. Law Rep.* 1997. V. 27. № 6. P. 10279–10285.
64. *Egilman D., Kim J., Biklen M.* Proving causation: the use and abuse of medical and scientific evidence inside the courtroom – an epidemiologist’s critique of the judicial interpretation of the Daubert ruling // *Food Drug Law J.* 2003. V. 58. № 2. P. 223–250.
65. *Hollingsworth J.G., Lasker E.G.* The Case against differential diagnosis: Daubert, medical causation. Testimony, and the scientific method // *J. Health Law.* 2004. V. 37. № 1. P. 85–111.
66. *Guzelian P.S., Victoroff M.S., Halmes N.C. et al.* Evidence-based toxicology: a comprehensive framework for causation // *Hum. Exp. Toxicol.* 2005. V. 24. № 4. P. 161–201. <https://doi.org/10.1191/0960327105ht517oa>
67. *Lagiou P., Trichopoulos D., Adami H.O.* Concepts in Cancer Epidemiology and Etiology // *Textbook of Cancer Epidemiology.* 2nd ed. // Eds H.O. Adami, D. Hunter, D. Trichopoulos. New York etc.: Oxford University Press, 2008. P. 127–152. <https://doi.org/10.1093/acprof:oso/9780195311174.003.0006>
68. *Greenland S.* Relation of probability of causation to relative risk and doubling dose: a methodologic error that has become a social problem // *Am. J. Public Health.* 1999. V. 89. № 8. P. 1166–1169. <https://doi.org/10.2105/AJPH.89.8.1166>
69. *The Health Consequences of Smoking: A Report of the Surgeon General* Rockville, MD: Office of the Surgeon General, US Public Health Service, 2004. 910 p.
70. *Mandil A.* Causal Inference in Epidemiology. Lecture Presentation // High Institute of Public Health. University of Alexandria. <https://ru.scribd.com/doc/306778054/Causal-Inference-in-Epidemiology>. Presentation: www.pitt.edu/~super4/33011-34001/33971.ppt (дата обращения 06.06.2019).
71. *Bhopal R.S.* Concepts of Epidemiology: Integrated the ideas, theories, principles and methods of epidemiology. 3rd ed. Oxford: University Press, 2016. 442 p.
72. *Fedak K.M., Bernal A., Capshaw Z.A., Gross S.* Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology // *Emerg. Themes Epidemiol.* 2015. V. 12. Article 14. <https://doi.org/10.1186/s12982-015-0037-4>

73. *Weed D.L., Trock B.J.* Criticism and the growth of epidemiologic knowledge. (Re: "Popperian refutation in epidemiology") // *Am. J. Epidemiol.* 1986. V. 123. № 6. P. 1119–1121.
<https://doi.org/10.1093/oxfordjournals.aje.a114341>
74. *Bruce N., Pope D., Stanistreet D.* Quantitative Methods for Health Research. A Practical Interactive Guide to Epidemiology and Statistics. 2nd ed. Oxford: John Wiley & Sons, 2019. 545 p.
75. *Rosnow R.L., Rosenthal R.* Effect Sizes for experimenting psychologists // *Canad. J. Experim. Psychol.* 2003. V. 57. № 3. P. 221–237.
<https://doi.org/10.1037/h0087427>
76. *Hinkle D.E., Wiersma W., Jurs S.G.* Applied Statistics for the Behavioral Sciences. 5th ed. Boston: Houghton Mifflin, 2003. 756 p.
77. *Rothman K.J., Greenland S.* Causation and causal inference in epidemiology // *Am. J. Public Health.* 2005. V. 95. Suppl 1. P. S144–S150.
<https://doi.org/10.2105/AJPH.2004.059204>
78. *Gordis L.* Epidemiology. 5th ed. Philadelphia: Saunders, Elsevier Inc., 2014. 392 p.
79. *Merrill R.M.* Introduction to Epidemiology. 7th ed. Burlington: Jones & Bartlett Learning, 2017. 339 p.
80. *Семеновых Г.К., Новиков С.М., Семеновых Л.Н.* Анализ случаев заболеваний, обусловленных действием факторов среды обитания. Характеристика опасности для здоровья: Учеб. пособие. Вып. 4. М.: Изд-во Первого МГМУ им. И.М. Сеченова, 2011. 88 с. [*Semenovykh G.K., Novikov S.M., Semenykh L.N.* Analysis of cases caused by the action of environmental factors. Characteristics of health hazards: Textbook. allowance. Issue 4. М.: Publishing House Sechenov First Moscow State Medical University, 2011. 88 p. (In Russian)]
81. *Хасанова Г.Р., Назарова О.А., Зорина Л.М., Хакимов Н.М.* Эпидемиология: Учебно-методическое пособие для студентов, обучающихся по направлению деятельности 34.03.01 "Сестринское дело". Казань: КГМУ, 2017. 73 с. [*Khasanova G.R., Nazarova O.A., Zorina L.M., Khakimov N.M.* Epidemiology: A teaching study guide for students enrolled in the direction of activity 34.03.01 "Nursing". Kazan: KSMU, 2017. 73 p. (In Russian)]
82. *Friis R.H., Sellers T.A.* Epidemiology for Public Health Practice. 5th ed. Burlington: Jones & Bartlett Learning, 2014. 804 p.
83. *Susser M.* Judgement and causal inference: criteria in epidemiologic studies // *Am. J. Epidemiol.* 1977. V. 105. № 1. P. 1–15. Reprint: *Am. J. Epidemiol.* 1995. V. 141. № 8. P. 701–715.
<https://doi.org/10.1093/oxfordjournals.aje.a117491>
84. *Susser M.* Rules of inference in epidemiology // *Regul. Toxicol. Pharmacol.* 1986. V. 6. № 2. P. 116–128.
[https://doi.org/10.1016/0273-2300\(86\)90029-2](https://doi.org/10.1016/0273-2300(86)90029-2)
85. *Susser M.* The logic of Sir Karl Popper and the practice of epidemiology // *Am. J. Epidemiol.* 1986. V. 124. № 5. P. 711–718.
<https://doi.org/10.1093/oxfordjournals.aje.a114446>
86. *Susser M.* Glossary: causality in public health science // *J. Epidemiol. Community Health.* 2001. V. 55. P. 376–378.
<https://doi.org/10.1136/jech.55.6.376>
87. *Котеров А.Н., Жаркова Г.П., Бiryukov А.П.* Тандем радиационной эпидемиологии и радиобиологии для практики радиационной защиты // *Мед. радиология и радиац. безопасность.* 2010. Т. 55. № 4. С. 55–84. [*Koterov A.N., Zharkova G.P., Biryukov A.P.* Tandem of radiation epidemiology and radiobiology for practice and radiation protection // *Medits. Radiologii Radiat. Bezopasnost (Medical Radiology and Radiation Safety, Moscow).* 2010. V. 55. № 5. P. 48–73. (In Russian. Engl. abstr.)]
88. *Lanes S.F.* The logic of causal inference // *Causal Inference / Ed. K.J. Rothman.* Chestnut Hill, MS: Epidemiologic Resources, 1988. P. 59–75.
89. *Maclure M.* Refutation in Epidemiology: why else not? // *Causal Inference / Ed. K.J. Rothman.* Chestnut Hill, MS: Epidemiologic Resources, 1988. P. 131–138.
90. *Rothman K.J.* What is causation // *Epidemiology, an Introduction / Ed. K.J. Rothman.* New York: Oxford University Press, 2002. P. 8–24.
91. *Current Topics in Occupational Epidemiology / Ed. K.M. Venables.* New York: Oxford University Press, 2013. 270 p.
92. *Rothman K.J.* Inferring causal connection – habit, faith or logic? // *Causal inference / Ed. K.J. Rothman.* Epidemiology Resources Inc., Mass., USA, 1988. P. 3–12.
93. *Karhausen L.R.* The poverty of Popperian epidemiology // *Int. J. Epidemiol.* 1995. V. 24. № 5. P. 869–874.
<https://doi.org/10.1093/ije/24.5.869>
94. *Parascandola M.* Epidemiology: second-rate science? // *Public Health Rep.* 1998 V. 113. № 4. P. 312–320.
95. *Lea H.C.* A History of the Inquisition of the Middle Ages. New York: Harper & Brothers, 1887. 583 p. [*Лу Г.Ч.* Инквизиция. Происхождение и устройство // *Беммер Г.* Иезуиты. Ли Ч. Инквизиция. СПб.: ООО "Изд-во ПОЛИГОН", 1999. 1248 с.]
96. *Strom B.L.* Study designs available for pharmacoepidemiology studies // *Pharmacoepidemiology.* 3rd ed. // *Ed. B.L. Strom.* Baffins Lane, Chichester, West Sussex: John Wiley & Sons Ltd, 2000. P. 17–30.
97. USEPA 2005. Guidelines for Carcinogen Risk Assessment. EPA/630/P-03/001B. Washington, DC: Risk Assessment Forum. National Center for Environmental Assessment Office of Research and Development U.S. Environmental Protection Agency, 2005. 166 p.
98. *Howick J., Glasziou P., Aronson J.K.* The evolution of evidence hierarchies: what can Bradford Hill's 'guidelines for causation' contribute? // *J. R. Soc. Med.* 2009. V. 102. № 5. P. 186–194.
<https://doi.org/10.1258/jrsm.2009.090020>
99. *Davey Smith G.* Data dredging, bias, or confounding. They can all get you into the BMJ and the Friday papers // *Brit. Med. J.* 2002. V. 325. № 7378. P. 1437–1438.
<https://doi.org/10.1136/bmj.325.7378.1437>
100. IARC 1987. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Supplement 7, Overall

- Evaluations of Carcinogenicity: An Updating of IARC Monographs, V. 1 to 42. Lyon, 1987. 449 p.
101. *Goodman K.J., Phillips C.V.* Hill's criteria of causation // In: 'Encyclopedia of Statistics in Behavioral Science. V. 2 // Eds B.S. Everitt, D.C. Howell. Chichester: John Wiley & Sons, Ltd, 2005. P. 818–820.
 102. *Vandenbroucke J.P.* Observational research, randomised trials, and two views of medical science // *PLoS Med.* 2008. V. 5. № 3. P. Art. e67. 5 p. <https://doi.org/10.1371/journal.pmed.0050067>
 103. *Schlesselman J.J.* "Proof" of cause and effect in epidemiologic studies: criteria for judgment // *Prev. Med.* 1987. V. 16. № 2. P. 195–210. [https://doi.org/10.1016/0091-7435\(87\)90083-1](https://doi.org/10.1016/0091-7435(87)90083-1)
 104. *Weed D.L.* Epidemiologic evidence and causal inference // *Hematol. Oncol. Clin. North Am.* 2000. V. 14. № 4. P. 797–807. [https://doi.org/10.1016/S0889-8588\(05\)70312-9](https://doi.org/10.1016/S0889-8588(05)70312-9)
 105. *Becker R.A., Dellarco V., Seed J. et al.* Quantitative weight of evidence to assess confidence in potential modes of action // *Regul. Toxicol. Pharmacol.* 2017. V. 86. P. 205–220. <https://doi.org/10.1016/j.yrtph.2017.02.017>
 106. *Koterov A.N., Biryukov A.P.* The possibility of determining of anomalies and pathologies in the offspring of liquidators of Chernobyl accident by the non-radiation factors // *Int. J. Low Radiat. (Paris)*. 2011. V. 8. № 4. P. 256–312. <https://doi.org/10.1504/IJLR.2011.046529>
 107. *Котеров А.Н., Бирюков А.П.* Дети ликвидаторов аварии на Чернобыльской атомной электростанции. 1. Оценка принципиальной возможности зарегистрировать радиационные эффекты // *Мед. радиология и радиац. безопасность*. 2012. Т. 57. № 1. С. 58–79. [*Koterov A.N., Biryukov A.P.* The offspring of liquidators of Chernobyl Atomic Power Station accident. 1. The estimation of the basic opportunity to register of radiation effect // *Medits. Radiologiya Radiat. Bezopasnost (Medical Radiology and Radiation Safety, Moscow)*. 2012. V. 57. № 1. P. 58–79. (In Russian. Engl. abstr.)]
 108. *Котеров А.Н., Бирюков А.П.* Дети участников ликвидации последствий аварии на Чернобыльской атомной электростанции. 2. Частота отклонений и патологий и их связь с нерадиационными факторами // *Мед. радиология и радиац. безопасность*. 2012. Т. 57. № 2. С. 51–77. [*Koterov A.N., Biryukov A.P.* The offspring of liquidators of Chernobyl Atomic Power Station accident. 2. The frequency of anomalies and pathologies and its connection to non-radiation factors // *Medits. Radiologiya Radiat. Bezopasnost (Medical Radiology and Radiation Safety, Moscow)*. 2012. V. 57. № 2. P. 51–77. (In Russian. Engl. abstr.)]
 109. *Gotzsche P.C.* Deadly Medicines and Organised Crime. How Big Pharma has Corrupted Healthcare. London: Radcliffe Publishing, 2013. 310 p. [*Gemue П.* Смертельно опасные лекарства и организованная преступность: как большая фарма коррумпировала здравоохранение: Пер. с англ. Л.Е. Зиганшиной. М.: Изд-во "Э", 2016. 464 с.]
 110. *Fox G.A.* Practical causal inference for ecoepidemiologists // *J. Toxicol. Environ. Health*. 1991. V. 33. № 4. P. 359–273. <https://doi.org/10.1080/15287399109531535>
 111. *Rothman K.J., Greenland S., Poole C., Lash T.L.* Causation and causal inference // *Modern Epidemiology* / Eds K.J. Rothman, S. Greenland, T.L. Lash. 3rd ed. Philadelphia (PA): Wolters Kluwer, 2008. P. 5–31.
 112. *Armenian H.K., Lilienfeld A.M.* The distribution of incubation periods of neoplastic diseases // *Am. J. Epidemiol.* 1974. V. 99. № 2. P. 92–100. <https://doi.org/10.1093/oxfordjournals.aje.a121599>
 113. *Hallquist A., Hardell L., Degerman A. et al.* Medical diagnostic and therapeutic ionizing radiation and the risk for thyroid cancer: a case-control study // *Eur. J. Cancer Prev.* 1994. V. 3. № 3. P. 259–267. <https://doi.org/10.1097/00008469-199403030-00004>
 114. *Wingren G., Hallquist A., Hardell L.* Diagnostic X-ray exposure and female papillary thyroid cancer: a pooled analysis of two Swedish studies // *Eur. J. Cancer Prev.* 1997. V. 6. № 6. P. 550–556. <https://doi.org/10.1097/00008469-199712000-0001>
 115. *Memon A., Godward S., Williams D. et al.* Dental x-rays and the risk of thyroid cancer: a case-control study // *Acta Oncol.* 2010. V. 49. № 4. P. 447–453. <https://doi.org/10.3109/02841861003705778>
 116. *Hallquist A., Jansson P.* Self-reported diagnostic X-ray investigation and data from medical records in case-control studies on thyroid cancer: evidence of recall bias? // *Eur. J. Cancer Prev.* 2005. V. 14. № 3. P. 271–276. <https://doi.org/10.1097/00008469-200506000-00012>
 117. *Zeeb H., Merzenich H., Wicke H., Blettner M.* Radiation Epidemiology // *Schottenfeld and Fraumeni Cancer Epidemiology and Prevention*. 4th ed. / Eds M.J. Thun et al. New York: Oxford University Press. Printed by Sheridan Books, Inc., USA. 2018. P. 2003–2037.
 118. *Bonita R., Beaglehole R., Kjellstrom T.* Basic epidemiology. 2nd ed. World Health Organization, 2006. 212 p.
 119. *Merrill R.M., Frankenfeld C.L., Freeborne N., Mink M.* Behavioral Epidemiology. Principles and Applications. Burlington: Jones & Bartlett Learning, LLC, 2016. 298 p.
 120. *Lagiou P, Adami H.O., Trichopoulos D.* Causality in cancer epidemiology // *Eur. J. Epidemiol.* 2005. V. 20. № 7. P. 565–574. <https://doi.org/10.1007/s10654-005-7968-y>
 121. *Fedeli U., Girardi P., Gardiman G. et al.* Mortality from liver angiosarcoma, hepatocellular carcinoma, and cirrhosis among vinyl chloride workers // *Am. J. Ind. Med.* 2019. V. 62. № 1. P. 14–20. <https://doi.org/10.1002/ajim.22922>
 122. *Vineis P., Kriebel D.* Causal models in epidemiology: past inheritance and genetic future // *Environ. Health*. 2006. V. 5. Art. 21. 10 p. <https://doi.org/10.1186/1476-069X-5-21>
 123. *Rogan W.J.* PCBs and cola-colored babies: Japan, 1968, and Taiwan, 1979 // *Teratology*. 1982. V. 26. № 3. P. 259–261. <https://doi.org/10.1002/tera.1420260307>

124. *Ha M., Lee S.Y., Hwang S.S. et al.* Evaluation report on the causal association between humidifier disinfectants and lung injury // *Epidemiol. Health.* 2016. V. 38. Art. e2016037.
<https://doi.org/10.4178/epih.e2016037>
125. *Mayer A., Hamzeh N.* Beryllium and other metal-induced lung disease // *Curr. Opin. Pulm. Med.* 2015. V. 21. № 2. P. 178–184.
<https://doi.org/10.1097/MCP.0000000000000140>
126. *Churg A., Wright J.L., Vedal S.* Fiber burden and patterns of asbestos-related disease in chrysotile miners and millers // *Am. Rev. Respir. Dis.* 1993. V. 148. № 1. P. 25–31.
<https://doi.org/10.1164/ajrccm/148.1.25>
127. *Weiss N.S.* Can the ‘specificity’ of an association be rehabilitated as a basis for supporting a causal hypothesis? // *Epidemiology.* 2002. V. 13. № 1. P. 6–8.
128. UNSCEAR 2000. Report to the General Assembly, with Scientific Annex. Annex F. DNA repair and mutagenesis. New York, 2000. P. 1–72.
129. *Shakir S.A., Layton D.* Causal association in pharmacovigilance and pharmacoepidemiology: thoughts on the application of the Austin Bradford-Hill criteria // *Drug Saf.* 2002. V. 25. № 6. P. 467–471.
<https://doi.org/10.2165/00002018-200225060-00012>
130. *Glynn J.R.* A question of attribution // *Lancet.* 1993. V. 342. № 8870. P. 530–532.
[https://doi.org/10.1016/0140-6736\(93\)91651-2](https://doi.org/10.1016/0140-6736(93)91651-2)
131. UNSCEAR 2006. Report to the General Assembly, with Scientific Annexes. Annex A. Epidemiological studies of radiation and cancer. New York: United Nations, 2008. P. 17–322.
132. *Gage S.H., Munafo M.R., Davey Smith G.* Causal inference in developmental origins of health and disease (DOHaD) research // *Annu. Rev. Psychol.* 2016. V. 67. P. 567–585.
<https://doi.org/10.1146/annurev-psych-122414-033352>
133. *Davey Smith G.* Assessing intrauterine influences on offspring health outcomes: Can epidemiological studies yield robust findings? // *Basic Clin. Pharmacol. Toxicol.* 2008. V. 102. № 2. P. 245–256.
<https://doi.org/10.1111/j.1742-7843.2007.00191.x>
134. *Davey Smith G., Phillips A.N., Neaton J.D.* Smoking as ‘independent’ risk factor for suicide: illustration of an artifact from observational epidemiology? // *Lancet.* 1992. V. 340. № 8821. P. 709–712.
[https://doi.org/10.1016/0140-6736\(92\)92242-8](https://doi.org/10.1016/0140-6736(92)92242-8)
135. *Aschengrau A., Seage G.R., III.* *Epidemiology in Public Health.* 3rd ed. Burlington: Jones & Bartlett Learning, LLC, 2014. 596 p.
136. *Gay J.* *Clinical Epidemiology & Evidence-Based Medicine Glossary: Terminology Specific to Epidemiology,* 2005.
<http://people.vetmed.wsu.edu/jmgay/courses/GlossEpiTerminology.htm> (address data 11.10.2020).
137. Smoking and reproductive life. The impact of smoking on sexual, reproductive and child health // Eds D. Carter D., N. Nathanson, C. Seddon et al. *British Medical Association. Board of Science and Education & Tobacco Control Resource Centre,* 2004.
https://www.rauchfrei-info.de/fileadmin/main/data/Dokumente/Smoking_ReproductiveLife.pdf (address data 11.10.2020).
138. UNSCEAR 2000. Report to the General Assembly, with Scientific Annex G. Biological effects at low radiation doses. New York, 2000. P. 73–175.
139. *Boice J.D. Jr.* Radiation epidemiology and recent paediatric computed tomography studies // *Ann. ICRP.* 2015. V. 44. № 1. Suppl. P. 236–248.
<https://doi.org/10.1177/0146645315575877>
140. *Meek M.E., Palermo C.M., Bachman A.N. et al.* Mode of action human relevance (species concordance) framework: Evolution of the Bradford Hill considerations and comparative analysis of weight of evidence // *J. Appl. Toxicol.* 2014. V. 34. № 6. P. 595–606.
<https://doi.org/10.1002/jat.2984>
141. *Guyatt G.H., Oxman A.D., Kunz R. et al.* (GRADE Working Group). Incorporating considerations of resources use into grading recommendations // *Brit. Med. J.* 2008. V. 336. № 7654. P. 1170–1173.
<https://doi.org/10.1136/bmj.39504.506319.80>
142. *Weed D.L.* Interpreting epidemiological evidence: how meta-analysis and causal inference methods are related // *Int. J. Epidemiol.* 2000. V. 29. № 3. P. 387–390.
<https://doi.org/10.1093/intjepid/29.3.387>
143. *Xu C., Cheng L.L., Liu Y. et al.* Protocol registration or development may benefit the design, conduct and reporting of dose-response meta-analysis: empirical evidence from a literature survey // *BMC Med. Res. Methodol.* 2019. V. 19. № 1. P. 78.
<https://doi.org/10.1186/s12874-019-0715-y>
144. USEPA 2006. A Framework for Assessing Health Risks of Environmental Exposures to Children. EPA/600/R-05/093F. Washington, DC: National Center for Environmental Assessment Office of Research and Development U.S. Environmental Protection Agency, 2006. 145 p.
145. *Lanes S.F., Poole C.* “Truth in packaging?” The unwrapping of epidemiologic research // *J. Occup. Med.* 1984. V. 26. № 8. P. 571–574.
146. *Liang W., Chikritzhs T.* Observational research on alcohol use and chronic disease outcome: new approaches to counter biases // *Sci. World J.* 2013. Article ID 860915. 14 p.
<https://doi.org/10.1155/2013/860915>
147. *Murata M., Miyake T., Inoue Y. et al.* Life-style and other characteristics of radiation workers at nuclear facilities in Japan: base-line data of a questionnaire survey // *J. Epidemiol.* 2002. V. 12. № 4. P. 310–319.
<https://doi.org/10.2188/jea.12.310>
148. USEPA 1998. Guidelines for Carcinogen Risk Assessment. EPA/630/R-95/002F. –Washington, DC: Risk Assessment Forum, 1998. 188 p.
149. *Ecological Causal Assessment / Eds S.B. Norton, S.M. Cormier, Glenn W. Suter II.* U.S. Environmental Protection Agency. Cincinnati, OH, USA: CRC Press, 2015. 497 p.
150. *Kaufman J.S., Poole C.* Looking back on ‘causal thinking in the health sciences’ // *Annu. Rev. Public Health.*

2000. V. 21. P. 101–119.
<https://doi.org/10.1146/annurev.publhealth.21.1.101>
151. *Котеров А.Н.* От очень малых до очень больших доз радиации: новые данные по установлению диапазонов и их экспериментально-эпидемиологические обоснования // Мед. радиология и радиац. безопасность. 2013. Т. 58. № 2. С. 5–21. [*Kotero A.N.* From very low to very large doses of radiation: new data on ranges definitions and its experimental and epidemiological basing // *Medit. Radiologia Radiat. Bezopasnost (Medical Radiology and Radiation Safety, Moscow)*. 2013. V. 58. № 2. P. 5–21. (In Russian. Engl. abstr.)]
152. *Luckey T.D.* Documented optimum and threshold for ionising radiation // *Int. J. Nuclear Law*. 2007. V. 1. № 4. P. 378–409.
<https://doi.org/10.1504/IJNUCL.2007.014806>
153. *Dauer L.T., Brooks A.L., Hoel D.G. et al.* Review and evaluation of updated researches on the health effects associated with low-dose ionizing radiation // *Radiat. Prot. Dosim.* 2010. 140. № 2. P. 103–136.
<https://doi.org/10.1093/rpd/ncq141>
154. UNSCEAR 2012. Report to the General Assembly, with Scientific Annexes. Annex A. Attributing health effects to ionizing radiation exposure and inferring risks. New York: United Nations, 2015. 86 p.
155. *Boice J.D. Jr.* The linear nonthreshold (LNT) model as used in radiation protection: an NCRP update // *Int. J. Radiat. Biol.* 2017. V. 93. №10. P. 1079–1092.
<https://doi.org/10.1080/09553002.2017.1328750>
156. NCRP Report No 136. Evaluation of the linear-non-threshold dose-response model for ionizing radiation. National Council on Radiation Protection and Measurements, 2001. 263 p.
157. *Hofler M.* The Bradford Hill considerations on causality: a counterfactual perspective // *Emerg. Themes Epidemiol.* 2005. V. 2. №. 11. 9 p.
<https://doi.org/10.1186/1742-7622-2-11>
158. *Котеров А.Н.* Малые дозы радиации: Факты и мифы. Основные понятия и нестабильность генома. М.: Изд-во “ФМБЦ им. А.И. Бурназяна ФМБА России”, 2010. 283 с. [*Kotero A.N.* Low Dose of Radiation: the Facts and Myths. The Basic Concepts and Genomic Instability. Moscow, Publ. by “FMBC by A.I. Burnazjan FMBA of Russia”, 2010. 283 p. (In Russian. Engl. abstr.)]

Causal Criteria in Medical and Biological Disciplines: History, Essence and Radiation Aspect. Report 3, Part 1: First Five Hill’s Criteria: Use and Limitations

A. N. Kotero^{a, #}

^a A. I. Burnasyan Federal Medical Biophysical Center of Federal Medical Biological Agency, Moscow, Russia

[#]E-mail: govorilga@inbox.ru

Report 3 is devoted to the history, nature and limitations of the epidemiological criteria of causation (‘Hill’s criteria’). Based on material from the original publications of leading researchers of causality (A.B. Hill., M.W. Susser, K. Rothman and others; 1950s–2019), from dozens of modern textbooks on epidemiology and carcinogenesis, from documents of international organizations (UNSCEAR, BEIR, USEPA, IARC, etc.), as well as from many other sources, in part 1 of the report, the first five Hill criteria are considered: strength of association, consistency of association, specificity, temporality and biological gradient (dose-response relationship). The theoretical and practical aspects for each criterion are presented: history of appearance, terminology, philosophical and epidemiological essence, applicability in various disciplines and limitations. For the strength association criteria the gradations (ordinal scales) are presented in terms of relative risk, odds ratio and Pearson’s (Spearman’s) correlation coefficient. Factual examples are provided for each of the criteria, including data from Radiation Epidemiology and Radiation Medicine.

Keywords: criteria for causality, the Hill criteria, strength, consistency, specificity, temporality, biological gradient