**ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ**

**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ**

**«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**

**Факультет государственного и муниципального управления**

**Кафедра управления и экономики здравоохранения**

**ДИССЕРТАЦИЯ НА СОИСКАНИЕ СТЕПЕНИ МАГИСТРА**

на тему:
**Возможные смещения при оценке эффективности лечения методом доказательной медицины в серийных исследованиях.**

Студентка группы № 762

Ракина Е.А.

Научный руководитель:

проф. д.м.н. Власов В.В.

Рецензент:
проф. д.м.н. Линденбратен А.Л.

**Москва, 2014г.**

Содержание

Введение 3

Глава 1. Несовершенства в применении научно обоснованных методов в доказательной медицине 11

1.1. Публикационное смещение 11

1.2. Фиксация побочных эффектов вмешательств 20

1.3. Лишние вмешательства 21

1.4. Унификация вмешательств 22

1.5. Вероятность ложноположительного результата испытания 24

Глава 2. Насколько часто положительный результат оценки эффективности вмешательства оказывается ложным 34

2.1. Анализ серийных клинических испытаний 34

2.2. Основные выводы 42

Глава 3. Практическое применение научно обоснованных методов с учетом возможных смещений 46

Заключение 54

Литература 58

# Введение

В настоящее время мировое медицинское сообщество рассматривает доказательную медицину не только как набор методов, помогающих врачу в принятии клинических решений или прогнозировании исходов. Во многих развитых странах принципы научной обоснованности являются основой для принятия управленческих и экономических решений в здравоохранении. Такие авторитетные организации как Управление по надзору за качеством пищевых продуктов и медикаментов (Food and Drug Administration, FDA, США) [[50](#_ENREF_50)], Национальный институт здоровья и клинического совершенствования Великобритании (National Institute for Health and Care Excellence, NICE) [[53](#_ENREF_53)], Институт качества и эффективности в здравоохранении в Германии (Institute for Quality and Efficiency in Health Care, IQWIG) [[51](#_ENREF_51)] в своих решениях относительно разрешения оборота лекарств, в оценке технологий здравоохранения, в том числе вопросов финансирования (предоставления) их основываются на принципах доказательной медицины и требуют подтверждения эффективности рассматриваемых вмешательств строгими научными методами. Аналогичные институты организуются и в других развитых странах [12].

Специалисты уже много лет говорят о необходимости создания подобных структур в России, а также о важности развития доказательной медицины для всей системы здравоохранения. Еще в 2005 году на Международном совещании на базе ММА им. И.М. Сеченова его участники сочли необходимым «просить Министерство здравоохранения и социального развития РФ рассмотреть вопрос о возможности создания структуры по оценке медицинских технологий». Также отмечалось, что, используя принципы доказательной медицины, можно более рационально расходовать ресурсы общественного здравоохранения, то есть создавать более эффективную систему здравоохранения по сравнению с существующей [7].

Подходы к медицинской практике и общественному здоровью, основанные на научных доказательствах (доказательная практика) не имеют рациональной альтернативы. Тем не менее, они имеют свои слабые места. Эти слабости не означают ущербности по сравнению с другими подходами. Они лишь означают, что наличными средствами надежность решений, обосновываемых научными данными, может быть относительно низкой.

Для некоторых, основных случаев обоснования решений, уже найдены достаточно эффективные методы работы. Так, например, для случая отсутствия больших доброкачественных испытаний эффективности и безопасности вмешательств была разработана методика мета-анализа. Результат каждого отдельного исследования может отличаться от истинного результата вмешательства в силу случайных или систематических ошибок при проведении исследования. Объединяя в мета-анализ результаты нескольких независимых исследований, мы снижаем влияние случайных ошибок, а в некоторых случаях и систематических смещений.

Согласно современным представлениям о доказательной медицине мета-анализ нескольких исследований считается медицинским утверждением, обладающим наибольшей доказательной силой по сравнению с отдельными исследованиями или экспертными мнениями. Но в силу ряда причин проведение мета-анализа также имеет свои ограничения. Во-первых, разные исследования одного и того же вмешательства могут опираться на разные конечные точки, поэтому их результаты будут несопоставимы между собой. Во-вторых, не для всех вмешательств можно найти даже два исследования, которые бы можно было бы включить в один мета-анализ. Такая проблема остро стоит для редких вмешательств, например, лечения орфанных заболеваний, когда количество потенциальных участников исследований невелико в силу небольшого количества пациентов. Кроме того это актуально и для новых препаратов и вмешательств, с которыми еще не провели достаточного для мета-анализа количества исследований. В-третьих, присутствующее в медицинской периодике публикационное смещение может сказаться на результатах мета-анализа и привести к их отклонению от истинного значения эффекта вмешательства.

Публикационное смещение, кроме влияния на результаты мета-анализов, также само по себе является одним из изъянов методов, применяемых в доказательной медицине для научного обоснования принятия решений. Также как несовершенство методов, используемых доказательной медициной, можно рассматривать проблемы, связанные с недостаточной фиксацией и преуменьшением побочных явлений в исследованиях. При современных подходах к проведению исследований, анализу их результату и алгоритмах принятия решений (учитывая бенефициаров этих решений) проявляется тенденция к рекомендации лишних вмешательств, в которых у пациента не было необходимости. Подсчет размера побочных эффектов от совершения необоснованных вмешательств по сравнению с невмешательством затруднен. Публикационное смещение традиционно учитывается при составлении систематических обзоров и мета-анализа с помощью воронкообразного графика, который позволяет с некоторой долей субъективности, вынести суждение о наличии или отсутствии публикационного смещения по данной проблеме, а также, иногда о степени его выраженности.

Существует сложность и в самих вмешательствах. Если вмешательство является чем-то более сложным, чем прием пациентом таблетки, то провести вмешательство можно по-разному. Замечено, что разные исследователи, используя одни и те же термины для описания вмешательств в публикациях, на самом деле могут производить с пациентом достаточно отличающиеся друг от друга действия. Сравнение результатов двух таких различных исследований может быть некорректным.

В данной работе также рассматривается еще одно возможное слабое место в применении принципов доказательной медицины. При проведении серийных исследований одного вмешательства (например, нового лекарственного препарата) возможна ситуация, когда «положительный» результат исследования, то есть статистически значимые различия между контрольной и экспериментальной группами будет являться не истинным значением эффекта, оказываемого данным препаратом, а следствием ошибки первого рода. Чем больше исследований проведено, тем больше вероятность, что хотя бы одно из них будет результатом этой ошибки. Как описывалось выше, в случаях проведения нескольких сходных исследований одного и того же вмешательства, в последствии с ними можно произвести мета-анализ, который сможет продемонстрировать отсутствие статистически значимых различий по суммарному результату нескольких исследований. Но, если исследования одного и того же препарата проводить при разных заболеваниях или состояниях, или на разных популяциях, не поддающихся сравнению, или использовать для измерения эффекта вмешательства разные конечные точки, проведение мета-анализа исследований окажется невозможным. В таком случае исследование, которое изначально являлось результатом ошибки первого рода, может стать доказательной основой для подтверждения несуществующей в реальности эффективности вмешательства.

При принятии организационных и экономических решений в области здравоохранения, опираясь на строгие научные методы, необходимо понимать и учитывать их возможные слабые стороны. Поэтому изучение возможных несовершенств, которые могут возникнуть при принятии решений, основанных на строгих научных методах, применяемых в доказательной медицине, на уровне систем здравоохранения является важной составляющей организации здравоохранения. Организационные и экономические решения, принятые под влиянием смещенных результатов исследований, могут приводить к нерациональному расходованию средств. Ограниченные средства системы здравоохранения могут быть выделены на предоставление неэффективного медицинского вмешательства, нередко дорогостоящего, в ущерб эффективным и необходимым медицинским вмешательствам. На основе публикационного смещения и ложноположительных результатов клинических испытаний число медицинских вмешательств, которые будут рассматриваться как вмешательства с доказанной эффективностью, может постоянно расти, даже если не будет обнаруживаться ни одного нового действительно эффективного вмешательства.

Целью данного исследования является анализ возможных смещений при оценке эффективности медицинских вмешательств, выявляемых в серийных исследованиях, а также предложение путей решения для важнейших проблем, выявленных в ходе анализа.

Работа над данным исследованием предполагает решение следующих задач:

1. Изучить обнаруженные несовершенства в применении научно обоснованных методов для оценки эффективности медицинских вмешательств, обсуждаемых в публикациях;

2. Сформулировать алгоритмы для вычисления вероятности ложноположительного результата для серийных исследований, не поддающихся объединению для проведения мета-анализа;

3. Провести анализ существующих серийных исследований на предмет вероятности, что показанный в исследованиях положительный результат является ложноположительным.

4. Сформулировать рекомендации по учету возможных смещений при оценке эффективности медицинских вмешательств в серийных исследованиях для более обоснованного принятия решений о рациональном расходовании средств бюджета здравоохранения.

Объектом данного исследования являются рандомизированные контролируемые испытания новых препаратов, проводимые сериями в различных областях медицинского применения.

Предметом данного исследования являются заложенные в серийных испытаниях величины допустимых ошибок первого и второго рода, результаты отдельных испытаний в серии, вероятность получения ложноположительного результата, говорящего об эффективности исследуемого вмешательства, при проведении серийных испытаний.

Теоретико-методологической основой данного исследования послужили труды по доказательной медицине, эпидемиологии, статистическому анализу в медицине таких ученых как Роберт и Сюзанна Флетчеры и Эдвард Вагнер, В. Власов, С. Гланц, О. Реброва, а также публикации Дж. Иоаннидиса о имеющихся несовершенствах в применении научно обоснованных методов оценки эффективности медицинских вмешательств.

В качестве эмпирической базы исследования выбраны регистры клинических испытаний, публикации о результатах зарегистрированных в регистрах испытаний или же данные о том, что такие результаты не были опубликованы, данные о регистрации новых лекарственных препаратов.

В качестве гипотезы данного исследования взято предположение, что при проведении рандомизированных клинических испытаний новых препаратов в разных областях медицины, результаты которых не могут быть подвергнуты мета-анализу, данные об эффективности препарата в одной из областей применения могут быть результатом статистической ошибки первого рода. Подобные данные, представленные независимо от других исследований в серии, не показавших эффективности, могут послужить сильным доказательством для дальнейшего применения препарата в данной области. А также могут стать основой для принятия ошибочных экономических и организационных решений об оплате подобного лечения и о предпочтительном выборе данного лечения по сравнению с лечением другим препаратом или отсутствием лечения.

Научной новизной исследования является расчет вероятности ложноположительного результата исследования в серии исследований, основанный на заложенных уровнях ошибок первого и второго рода, числе исследований данного препарата, проведенных на данный момент, количестве исследований, показавших эффективность препарата, вне зависимости от медицинской сферы проведения отдельных испытаний препарата. В данной работе были проанализированы 70 исследований, зарегистрированных в регистрах клинических испытаний, проводившихся в связи с изучением пяти препаратов, одобренных FDA. Вычислены вероятности ложноположительных результатов среди исследований, показавших статистически значимые различия.

Практическая значимость работы состоит в формулировании предложений и рекомендаций по учету существующих слабых мест в доказательствах эффективности медицинских вмешательств. Правильный учет и анализ слабых и сильных сторон доказательств критически важен для принятия решений в области организации здравоохранения о закупках и предоставлении лекарственных препаратов, решений об оснащении учреждений здравоохранения теми или иными видами оборудования, принятия решений о включении исследований как профилактических или скрининговых. Именно от таких решений зависит эффективность и рациональность расходования средств здравоохранения. А поскольку ресурсы, в том числе денежные, имеющиеся в распоряжении системы здравоохранения, всегда ограничены, возможность правильно оценить силу доказательств и избежать трат на лечение, доказательство эффективности которого являются следствием ложноположительного результата, имеет важное организационное и экономическое значение для здравоохранения.

# Глава 1. Несовершенства в применении научно обоснованных методов в доказательной медицине

## 1.1. Публикационное смещение

За несколько десятилетий существования доказательной медицины как самостоятельного направления науки был выявлен ряд несовершенств, которые имеют научно обоснованные методы, применяемые в доказательной медицине для оценки эффективности вмешательств, и для некоторых из них были предложены решения.

Публикационное смещение является одним из важных явлений, оказывающих негативное влияние на правильность научных выводов. Это явление наблюдается не только в области доказательной медицины, но и в некоторых других экспериментальных науках. Проблема публикационного смещения состоит в том, что не все исследования, которые проводятся в области медицины и здравоохранения, в итоге оказываются опубликованными и доступными для дальнейшего анализа другими исследователями. Вероятность того, что некое исследование окажется неопубликованным или недоступным для других исследователей, не является случайной, а зависит от природы исследования и его результатов [[42](#_ENREF_42)]. Поскольку основным источником информации для медицинского сообщества являются именно публикации исследований, важно иметь возможность оценить наличие и степень публикационного смещения при сборе информации. Уменьшение публикационного смещения оказывается в интересах всего медицинского сообщества.

Публикационное смещение может возникать в следствие разных причин. Принято выделять несколько типов публикационного смещения [[42](#_ENREF_42)].

1. Смещение на этапе публикации. Будут опубликованы результаты исследования или не будут, может зависеть от результатов этого исследования.

2. Смещение во времени публикации. Время, за которое исследование будет полностью завершено, описано и опубликовано, может зависеть от результатов исследования.

3. Смещение в результате множественных публикаций. От результатов исследования может зависеть, будет ли оно опубликовано однократно или многократно.

4. Смещение из-за места размещения публикации. В зависимости от результатов исследования оно может попадать в журналы и базы данных с разной степенью доступности и цитируемости.

5. Смещение в результате цитирования. Уровень цитирования исследования зависит от результатов исследования.

6. Языковое смещение. Публикация исследования на определенных языках может зависеть от результатов, показанных в исследовании.

7. Смещение в сообщении исходов. Избирательное сообщение одних исходов и умалчивание о других исходах может зависеть от результатов, полученных по тем и другим исходам в исследовании.

Причина появления публикационных смещений разных типов состоит в том, что «положительный» результат исследования, то есть данные, доказывающие, что исследуемое вмешательство оказывает какой-то статистически значимый эффект, рассматривается исследователями, редакторами, издателями и читателями научной литературы как более значимый по сравнению с «отрицательным» результатом. В связи с этим исследования, показывающие статистически значимый «положительный» результат имеют больше шансов быть опубликованными по сравнению с исследованиями, показывающими, что эффекта вмешательства не обнаружено. Также исследования, доказывающие эффективность вмешательства, публикуются быстрее, у них больше вероятность быть опубликованными на английском языке и в более авторитетном журнале. Кроме того исследования с «положительным» результатом чаще публикуются повторно по сравнению с исследованиями об отсутствии значимого эффекта. А поскольку доказанное наличие статистически значимого результата кажется исследователям более весомым, чем отсутствие статистически значимого результата, то и цитируют такие исследования чаще [[42](#_ENREF_42)].

Публикационное смещение может иметь большое значение для систематических обзоров и мета-анализов [10]. Создание систематических обзоров с мета-анализами призвано обобщить результаты всех исследований, проведенных в какой-то сфере (это может быть отдельное вмешательство или несколько вмешательств при каком-то состоянии). Подобное обобщение результатов многих исследований с применением статистических методов позволяет нивелировать случайные и некоторые систематические ошибки, присутствующие в отдельных исследованиях, и с помощью вычислений приблизиться к истинному результату о размере эффекта вмешательства. Но при наличии публикационного смещения может сложиться ситуация, когда исследователю, проводящему мета-анализ, будут доступны исследования, показывающие наличие эффекта, и не доступны исследования, показывающие отсутствие эффекта. Результат такого мета-анализа будет сдвинут в сторону большего размера эффекта, что приведет к ошибочным выводам.

Из-за возможного искажения общей картины в результате публикационного смещения, а также в силу того, что любое исследование на людях несет в себе определенный риск и возможный вред для участников, вопрос публикации результатов исследования имеет этический аспект [5; 9]. Научное медицинское сообщество сходится в том, что должны публиковаться результаты всех клинических испытаний. Однако на практике исследователи не всегда выполняют свои обязательства по публикации исследований, не всегда публикуют результаты в полном объеме, поэтому возникает необходимость в механизмах, которые бы противодействовали бы публикационному смещение или давали бы возможность оценить наличие такого смещения и степень его выраженность.

В качестве одного из способов для оценки публикационного смещения используется создание воронкообразного графика. Воронкообразный график представляет собой простую точечную диаграмму по одной из осей которой отложен размер эффекта, показанный в отдельных исследованиях, а по другой оси указан размер исследования или его точность. Обычно эффект откладывается по горизонтальной оси, а мера величины исследования по вертикальной оси. Поскольку точность исследования (в качестве ее меры может быть взята дисперсия или стандартная ошибка) должна повышаться по мере увеличения размера исследования, то можно отмерять размер исследования его точностью. Кроме того по мере увеличения размера исследования, оценка имеющегося эффекта от исследуемого вмешательства должна становиться точнее. А оценка эффекта в более мелких исследованиях должна быть разбросана шире друг относительно друга. Таким образом в нижней части диаграммы находятся точки, обозначающие небольшие исследования, достаточно далеко разбросанные. В верхней части диаграммы отмечены более крупные исследования, чьи точки лежат более кучно. Итоговый график похож на перевернутую воронку [[36](#_ENREF_36); [42](#_ENREF_42)].

В случае имеющегося публикационного смещения, причиной которого являются обнаруженные результаты исследования, часть точек, которые должны были бы быть на диаграмме, будут отсутствовать. И весь график приобретет асимметричный вид. Обычно асимметричность графика возникает за счет отсутствия части малых по размеру (или точности) исследований.

Хотя этот метод и позволил обнаружить значительные смещения и выявить несовершенства в проведенных мета-анализах, сам он также имеет ряд ограничений. Публикационное смещение разных типов (в том числе смещение из-за времени и места публикации) является только одной из возможных причин асимметричности воронкообразного графика. Кроме него причиной асимметричности могут быть ошибки в дизайне исследований, что чаще встречается в небольших исследованиях, ошибки анализа результата или даже подлог результатов. В некоторых случаях асимметрия является результатом различия в популяциях, исследованных в малых и в больших исследованиях. Небольшие по размеру исследования могли проводиться на пациентах с более тяжелой формой заболевания, в которых ожидаемо размер эффекта будет больше, чем среди пациентов с менее выраженными проявлениями болезни. Для некоторых статистических показателей, например отношения рисков, ожидаема некоторая асимметрия, поскольку их параметры соотносятся со стандартной ошибкой. Также асимметрия может быть результатом случайности, особенно, если количество исследований, включенных в мета-анализ невелико, поэтому проверку воронкообразным графиком имеет смысл проводить, только если в анализ включено не менее десяти разных исследований [[36](#_ENREF_36); [42](#_ENREF_42)].

Вторым важным ограничением применения оценки по воронкообразному графику является то, что решение о том, является ли график асимметричным или симметричным достаточно субъективно. Только в случаях сильно выраженного смещения можно твердо говорить о асимметричности. Чтобы снизить влияние субъективного суждения специалиста, используются дополнительные методы оценки асимметричности графика, например, линейная регрессия [[15](#_ENREF_15)].

Другим способом избежать публикационного смещения и повысить ответственность исследователей по своевременной публикации результатов проведенных клинических испытаний является регистрация исследований в регистрах клинических испытаний. На данный момент существуют национальные и международные регистры клинических испытаний. Крупнейшим регистром является <http://clinicaltrials.gov/>, являющийся службой Национальных институтов здоровья США [[41](#_ENREF_41)], позволяющий регистрацию из любой точки мира, а также допускающий ретроспективную регистрацию исследований [[1](#_ENREF_1)]. Существует регистр Европейского Союза [www.clinicaltrialsregister.eu](http://www.clinicaltrialsregister.eu) и международный регистр Всемирной организации здравоохранения <http://apps.who.int/trialsearch/> [[45](#_ENREF_45); [52](#_ENREF_52)]. В России существуют реестры выданных разрешений на клинические испытания медицинских изделий и клинические исследования лекарственных препаратов, которые для некоторых целей могут быть использованы аналогично зарубежным регистрам клинических испытаний [[2](#_ENREF_2); [3](#_ENREF_3)].

Первые законодательные требования о регистрации клинических испытаний в общедоступной базе данных были приняты Конгрессом США в 1997 году [[49](#_ENREF_49)]. Международный комитет редакторов медицинских журналов (ICMJE) в 2005 году рекомендовал принимать к публикации только те отчеты о клинических испытаниях, которые были зарегистрированы в открытых регистрах клинических испытаний до или во время включения первого пациента в исследование. Такие же требования остались и в современной версии Рекомендаций о проведении, сообщении, редактировании и публикации научных работ в медицинских журналах [[57](#_ENREF_57)]. В 2006 году Всемирная организация здравоохранения заявила о необходимости регистрировать все проводимые клинические испытания. А в 2008 году в Хельсинскую декларацию было включено требование: «Любое клиническое испытание должно быть зарегистрировано в базе данных, доступной для просмотра населением, перед набором первого субъекта» [[4](#_ENREF_4)].

Проспективная регистрация клинических испытаний должна привести к тому, что информацию о том, что то или иное испытание было проведено, было невозможно утаить вне зависимости от полученных результатов исследования. Теоретически, это должно уменьшить публикационное смещение. Однако, на практике оказывается, что даже при проведении регистрации с публикацией могут возникать трудности. Во-первых, регистрация клинического испытания не избавляет от смещения публикации во времени. Даже если в итоге все результаты клинических испытаний будут опубликованы, можно предположить, что испытания, в которых были получены статистически значимые отличия, будут опубликованы раньше, чем клинические испытания, не показавшие статистически значимого результата. Во-вторых, это не избавит от языкового смещения и смещения места публикации. Вероятно, что на английском языке и в более авторитетных журналах будет больше опубликовано испытаний, показавших «положительный» результат. В-третьих, даже если все медицинские журналы будут публиковать только своевременно зарегистрированные клинические испытания, это не значит, что часть зарегистрированных испытаний не окажется неопубликованной вовсе. Исследования показывают, что только 10% от всех клинических испытаний, зарегистрированных на ClinicalTrials.gov на декабрь 2012 года, и 15% от зарегистрированных и завершенных клинических испытаний имеют опубликованные результаты [[31](#_ENREF_31)].

Такой объем неопубликованных результатов клинических исследований вызывает беспокойство и озабоченность у специалистов. В связи с этим в январе 2013 года была запущена кампания AllTrials, целью которой является регистрация, публикация и доступность всех результатов всех проводимых клинических испытаний [[1](#_ENREF_1)].

Значительное публикационное смещение наблюдается среди исследований, темой которых являются методы лечения в области психиатрии. Так в 2008 году был проанализирован отчет FDA об испытаниях антидепрессантов и соответствующие публикации исследований. Было выявлено, что из исследований, в которых приняло участие более 12000 пациентов, опубликованными оказались лишь 37 исследований из 38, которые FDA сочло показывающими положительный результат. А из тех 36 исследований, которые были расценены FDA как исследования с отрицательным или спорным результатом, 22 не были опубликованы вовсе. 11 из 14 опубликованных исследований предоставляли информацию о тех исходах, которые показали положительный результат. Если судить по публикациям, то 94% исследований имели положительный результат. Если судить по сообщению FDA, то эффективность лечения показал только 51% испытаний. Благодаря этому публикационному смещению оцениваемый эффект от лечения увеличился на 32% [[24](#_ENREF_24)].

Значительное смещение во времени публикаций было обнаружено в исследованиях терапии ВИЧ. Было обнаружено, что как минимум 13 испытаний 6 антивирусных препаратов оставались не опубликованными не менее трех лет с момента окончания испытания. Эти исследования включали в себя не менее 3779 пациентов. По крайней мере 9 из этих 13 испытаний получили отрицательный предварительный или итоговый результат. Два больших исследования изопринозина, показавшие отрицательный результат, остались не опубликованными, тогда как результаты испытания, показавшего положительный результат, было опубликовано в авторитетном журнале с высоким импакт-фактором [[24](#_ENREF_24)].

Для организации и экономики здравоохранения публикационное смещение имеет важные отрицательные последствия. Под влиянием публикационного смещения все исследуемые в клинических испытаниях медицинские вмешательства оказываются более эффективными, чем они есть на самом деле. То есть осмысленность и эффективность денежных трат на предоставление новых лечебных, профилактических и диагностических вмешательств оказывается стабильно завышенной по сравнению с реальным положением дел. Аналогично новое оборудование, которое обычно является и более дорогим, чем существующее, благодаря публикационному смещению может быть подано как более эффективное, дающее более точную диагностику или больший лечебный эффект. Таким образом публикационное смещение может приводить к неверно принятым решениям о трате дополнительных денежных ресурсов системы здравоохранения на неэффективные вмешательства.

Как видно из приведенных примеров публикационного смещения, эта проблема затрагивает самые разные области здравоохранения. Решение о предоставлении или не предоставлении пациентам новых препаратов для лечения депрессии или ВИЧ, а также других частых заболеваний, может иметь значительные экономические последствия. Во-первых, новые препараты почти всегда оказываются дороже старых, поэтому решение предоставить новый препарат, который кажется более эффективным, будет стоить системе здравоохранения больших расходов. Во-вторых, если новый препарат окажется менее эффективным, чем использовавшийся прежде, то большее число пациентов окажется невылеченными и, соответственно, не смотря на увеличение расходов, бремя болезни для общества также возрастет.

Оценка публикационного смещения должна производится всегда, когда речь идет об оценке эффективности препарата и связанных с этим решениях будь то регистрация препарата для доступа на рынок, решение о предоставлении препарата за счет средств общественного здравоохранения, решение о включении препарата в перечни и стандарты лечения. Правильная оценка публикационного смещения значительно повышает рациональность и эффективность расходования средств здравоохранения.

## 1.2. Фиксация побочных эффектов вмешательств

Методы проведения клинических испытаний, которые потом используются доказательной медициной, имеют ограниченную силу в обнаружении побочных эффектов, в томи числе неблагоприятных. Поскольку большинство побочных эффектов достаточно редко встречаются, то они могут быть не обнаружены даже в достаточно больших рандомизированных клинических испытаниях. Из-за этого препарат может доказать свою эффективность в лечении, быть зарегистрированными и поступить в продажу, и только в последствии обнаружится важный побочный эффект, который может даже послужить причиной для запрета использования препарата. Иногда, вовремя не обнаруженный побочный эффект нового препарата является поводом для дискредитации научных методов, послуживших доказательной базой для принятия решения о регистрации препарата, как небезопасных.

Из-за редкости проявления побочных эффектов основным способом их изучения и выявления являются исследования типа сравнения с контролем. Они менее надежны чем контролируемые испытания или когортные исследования, но позволяют выявить более редкие нежелательные явления. Для хорошей работы по выявлению побочных эффектов требуется отлаженная система по регистрации всех осложнений. Но создание подобных регистров затруднено, поскольку сами врачи не всегда в них заинтересованы: немалая часть осложнений являются ятрогенными, и в интересах врача оказывается их сокрытие, а не регистрация. Нередко врач не может сам оценить, явилось ли осложнение следствием побочного действия лекарственного средства, на которое врач не мог повлиять, или следствием некачественной работы врача. В российской системе здравоохранение обвинение врача во всех проявившихся у пациента осложнениях является заметной проблемой, мешающей качественной регистрации побочных явлений от применяемых лекарственных средств. Возможным способом разрешения этой конфликтной для врачей ситуации может служить создание системы анонимной регистрации осложнений у пациентов, не предполагающей наказания для сообщившего[5].

## 1.3. Лишние вмешательства

Последние годы проблема лишних вмешательств в медицине стала оказывать значительное влияние на здравоохранение в целом. Современные исследователи предполагают, что в настоящее время немалое количество пациентов подвергается лишним диагностическим исследованиям и медицинским вмешательствам, которых можно было бы избежать. Эти вмешательства могут быть достаточно инвазивными и опасными, приносить пациенту вред, в некоторых случаях лишнее лечение может даже приводить к смерти пациента. Проблема лишних вмешательств и избыточной диагностики идет в ногу с развитием диагностического оборудования. Более точные и чувствительные приборы позволяют идентифицировать еще более мелкие изменения в организме человека, еще более ранние стадии заболеваний. Но не всегда оказывается возможным поставить точный критерий для отличия патологического состояния от нормального. Кроме того, оказывается, что не всегда начало лечения заболевания на более ранней стадии дает пациенту какие-то значимые преимущества, но может подвергать его дополнительным рискам.

Особенно остро проблема ненужного лечения стоит в области онкологии. С одной стороны это связано с опасностью онкологических заболеваний и, соответственно, стремлению выявить их как можно раньше и лечить радикально, не применяя выжидательной тактики. С другой стороны такой сдвиг диагностики связан по крайней мере частично с рентными установками онкологов. Большая часть применяемых в лечении онкологических заболеваний методов: хирургия, химиотерапия, облучение – сами по себе являются достаточно вредными или опасными [[16](#_ENREF_16)] и высокодоходными для провайдеров. Перед современной доказательной медициной стоит задача, с одной стороны, повысить чувствительность применяемых диагностических тестов, но одновременно, увеличить их специфичность. С проблемой гипердиагностики и лишнего лечения напрямую связан выбор тестов, применяемых для скрининга на различные заболевания. Сейчас происходит пересмотр целесообразности многих скрининговых исследований, в том числе таких распространенных как маммография [[14](#_ENREF_14)].

## 1.4. Унификация вмешательств

Корректно проведенные исследования позволяют на основании попавшей в исследование выборки сделать корректные вывод о всей популяции. А на основании принадлежности конкретного пациента к данной популяции, врач выбирает наиболее подходящее лечение. Поэтому при практическом применении принципов доказательной медицины остро стоит вопрос об обобщаемости полученных в исследованиях результатов. Врач может основываться в лечении пациента на результатах того или иного исследования в зависимости от того, насколько его пациент ближе к одной или другой исследованной выборке. При этом нередко из виду упускается само исследованное вмешательство. Если вмешательство заключается в простом приеме готового таблетированного препарата в указанной дозировке, с этим обычно не возникает трудностей. Но если исследуется какое-то более сложное вмешательство, некая операция, процедура, облучение, сложное введение препарата, диагностическое мероприятие, то одного указания названия оказывается недостаточно. Под одним и тем же названием процедуры или операции разные врачи – проводившие исследование и читающие его результаты – могут понимать значительно отличающиеся друг от друга действия. Методы проведения даже достаточно распространенных процедур могут отличаться от одной страны к другой, и даже в разных больницах одного государства. Описания состояния пациентов или результатов диагностики тоже могут быть достаточно субъективными. Исследования показывают, что даже в авторитетных рецензируемых журналах менее половины опубликованных испытаний содержат достаточно подробное и понятное описание проведенных вмешательств, которое позволяет корректно их воспроизвести другому врачу [[30](#_ENREF_30)].

Недостаточно подробные описания вмешательств могут иметь несколько последствий. На основании опубликованного исследования врач может неверно воспроизводить описанное вмешательство и тем самым предоставлять пациенту не оптимальную в его случае медицинскую помощь. В разных исследованиях под одним названием могут проводиться разные вмешательства, а последующее объединение их в систематических обзор или мета-анализ может привести к некорректным выводам. Подробное описание исследуемых вмешательство должно стать предметом внимания редакторов журналов, рецензентов и читателей. Возможным решением проблемы предоставления достаточно подробных описаний может статьи использование мультимедийных возможностей, например, размещение на сайтах журналов видеозаписей или схем проведения вмешательства.

Проблема недостаточной фиксации побочных эффектов исследуемых медицинских вмешательств, а также проблема неоправданных лишних вмешательств имеют выраженные негативные экономические последствия для системы здравоохранения. Заниженный уровень побочных эффектов от вмешательств приводит к недооценке их потенциальной опасности и неправильной оценке затрат и выгод. Остаются недооцененными экономические затраты на лечение побочных эффектов, на компенсацию потери трудоспособности, инвалидности, смертности. Нередко, истинный размер побочных эффектов превышает пользу от прямого действия исследуемого вмешательства, но эти последствия выясняются с опозданием.

Тенденция роста доли лишних вмешательств также остро актуальна для экономики здравоохранения. Постоянное развитие медицины само по себе ведет к увеличению расходов на здравоохранение, а лишние вмешательства делают растущие расходы все менее и менее эффективными. Кроме того рост доли лишних вмешательств приводит к увеличению побочных эффектов от этих вмешательств и финансовых затрат на их лечение, которых можно было бы избежать. Учитывая, что финансовые ресурсы здравоохранения всегда ограничены, трата денег на лишние вмешательства является большой нагрузкой на всю систему здравоохранения.

При принятии организационных и экономических решений в сфере здравоохранения следует учитывать, что доказательства эффективности вмешательства будут преувеличивать полезность вмешательства и преуменьшать побочные эффекты. Для корректировки оценки следует снижать расчетную эффективность и результативность рассматриваемого вмешательства и повышать потенциальную опасность и потенциальные расходы на лечение последствий побочных эффектов нового медицинского вмешательства.

## 1.5. Вероятность ложноположительного результата испытания

Доказательная медицина отличается от традиционной клинической практики тем, что предполагает рассматривать пациента в контексте различных больших популяций, к которым он может принадлежать, и давать рекомендации и прогнозы на основе оценки вероятностей событий. Доказательная медицина широко применяет экспериментальный подход, чтобы делать заключения, основываясь на строгих научных данных. Как и в других экспериментальных науках корректная интерпретация экспериментов и наблюдений основывается на статистической обработке результатов [8; 11].

Поскольку клинические испытания проводятся не на всей популяции пациентов, а на выборке, то для корректного обобщения результатов выборки на всю популяцию также всегда используются статистические методы [11].

Одним из основных экспериментальных методов доказательной медицины являются рандомизированные контролируемые испытания. Объектом данного исследования также являются рандомизированные контролируемые испытания, проведенные сериями для исследования новых препаратов. Общая схема проведения рандомизированных контролируемых испытаний может быть изложена кратко. Сначала определяется популяция пациентов, для которой должна быть проверена эффективность изучаемого вмешательства, например, приема нового препарата. Из этой популяции набирается некоторая выборка. Затем случайным образом выборка распределяется на группы, во многих случаях формируется только две группы, но их может быть и больше. Одной группе применяется исследуемое вмешательство, другая группа составляет контроль, и ей проводят лечение с помощью плацебо или с помощью другого лекарства, с которым желают сравнить исследуемый препарат. В конце испытания исследователи сравнивают результаты в группах по исходам, определенным как важные в начале исследования. Результат может оцениваться по наступлению каких-то событий, например, смерти, рецидива заболевания, инсульта, инфаркта, выздоровления, инвалидизации и других. В других исследованиях результат оценивается в степени изменении какого-то параметра, выражающего количественно, например, уровня артериального давления, уровня вирусной нагрузки, дозы принимаемого препарата и других показателей.

Сравнение результатов – важный этап клинического испытания, поскольку именно на основании сравнения будет сделан итоговый вывод, эффективен ли препарат по сравнению с плацебо (или эффективнее он другого препарата, если сравнение проводилось не с плацебо, а с другим видом лечения) или же препарат не показал никакой эффективности. Гипотезу о том, что препарат эффективнее плацебо, невозможно доказать подтверждающим эту гипотезу экспериментом. Поэтому доказательство строится от обратного. Сначала формулируется нулевая гипотеза, которая обычно противоположна исследовательской гипотезе. То есть в случае рандомизированных клинических испытаний нулевая гипотеза предполагает равенство сравниваемых групп. Если в результате проведенного исследования можно отклонить нулевую гипотезу, то принимается альтернативная гипотеза. В нашем случае это гипотеза состоит в том, что группы между собой различаются, и это различие не случайно, а отражает истинное различие между совокупностями пациентов, получавших испытываемое лечение или контрольное (активное или плацебо). Нулевую гипотезу проверяют с помощью статистических критериев [5; 10].

Статистические критерии не дают точного ответа о истинности или ошибочности нулевой гипотезы. По результатам применения статистического критерия можно говорить о некоторой вероятности, что тот или иной вывод верен. Сочетания истинности и ложности нулевой гипотезы и вероятности ее принятия или не принятия можно изобразить в виде четырехпольной таблицы.

Таблица 1. Варианты сочетаний результатов проверки нулевой гипотезы и истинным сравнением групп.

|  |  |  |
| --- | --- | --- |
|  |  | Наличие различий в действительности |
|  |  | Различия есть (нулевая гипотеза неверна) | Различий нет (нулевая гипотеза верна) |
| Результат применения статистического критерия | Различие выявлено (нулевая гипотеза отклоняется) | Истинно-положительный результат (1-β) | Ложно-положительный результатα(ошибка I рода) |
| Различий не выявлено (нулевая гипотеза принимается) | Ложно-отрицательный результатβ(ошибка II рода) | Истинно-отрицательный результат(1-α) |

Ошибкой первого рода (ошибка типа I, α-ошибка) называют ситуацию, когда нулевая гипотеза отклоняется ошибочно, то есть в исследовании обнаруживаются различия, которых на самом деле нет. Ошибкой второго рода (ошибкой типа II, β-ошибкой) называют обратную ситуацию, когда нулевая гипотеза ошибочно принимается как верная, то есть утверждается отсутствие различий, хотя на самом деле они существуют.

Ошибка первого рода также выражается как уровень статистической значимости p. Уровень допустимых в исследовании α и β ошибок устанавливается исследователем произвольно, в зависимости от природы исследуемого вмешательства и целей исследования. Традиционно пороговый уровень статистической значимости в медицинских исследованиях устанавливается 0,05. Уровень 0,01 считается высоко значимым. Для β-ошибки традиционный уровень составляет 0,2 [5; 10].

Традиция сравнивать уровень значимости с 0,05 и 0,01 отчасти связана с докомпьютерными вычислениями, когда сравнить значение критерия с табличными значениями было проще, чем рассчитать его точное значение. Сейчас, когда все статистические вычисления производятся доступными компьютерными средствами, нет необходимости указывать уровень статистической значимости как неравенство относительно пороговых величин, а корректно указывать его точное значение.

Кроме того существует два подхода к оценке уровня статистической значимости. В соответствии с подходом Фишера уровень статистической значимости отражает вероятность правильности или не правильности нулевой гипотезы, и исследователю следует оценивать точное значение p и делать выводы о том, насколько достоверно доказательство того, что нулевую гипотезу следует отклонить. Например, если значение p лежит в пределах от 0,1 до 0,9, то сомнений в верности нулевой гипотезы не должно быть. А если p<0,02, - это сильный показатель, что нулевая гипотеза не верна. В противоположность подходу Фишера, Нейман и Пирсон утверждали, что нельзя субъективно оценивать уровень полученного p как аргумент «за» или «против» нулевой гипотезы. Учитывая возможность ошибок первого и второго рода, они утверждали, что тесты, основывающиеся на теории вероятности, не могут быть доказательством опровержения гипотезы, а должны направлять исследователя в принятии решений. То есть согласно подходу Неймана и Пирсона следует принимать или отвергать нулевую гипотезу в соответствии с выбранным уровнем значимости, а не пытаться оценить значение p по силе его доказательства в пользу нулевой гипотезы в конкретном исследовании [[35](#_ENREF_35)].

Для корректного использования подхода Неймана-Пирсона следует достаточно точно формулировать альтернативную гипотезу: например, недостаточно просто предположить, что вмешательство эффективно, а следует также предположить и размер эффекта. Однако исследователи практически никогда не поступают так. Широкое распространение получила только та часть подхода Неймана-Пирсона, которая говорит о принятии или отвержении нулевой гипотезы по пороговому значению p=0,05 [[35](#_ENREF_35)].

К ложноположительным результатам может приводить не только сама вероятность ошибки первого рода, но и то значение, которое вкладывается в понятие уровня статистической значимости. Многие исследователи понимают p как вероятность справедливости нулевой гипотезы [6]. Но вероятность справедливости нулевой гипотезы зависит не только от уровня p, но и от самих проверяемых гипотез, какова среди них доля действительно ложных нулевых гипотез [[18](#_ENREF_18)]. Можно проиллюстрировать это утверждение простым расчетом. Предположим, что из тестируемых нами нулевых гипотез только 10% действительно ложны. А чувствительность критерия (1-β) предположим равной 50%, что часто встречается из-за небольшого размера исследований [6; [35](#_ENREF_35)]. Предположим, что мы проверим 1000 таких гипотез с критическим значением p<0,05. 100 из них будут действительно истинными, то есть нулевая гипотеза будет неверна. Но из-за ошибки второго рода равной 50% мы отвергнем только 50 из этих нулевых гипотез. Но кроме того, из оставшихся 900 исследований, для которых нулевая гипотеза верна, мы отвергнем 45 из-за ошибки первого рода, равной 5%. То есть из 95 исследований, результат которых мы сочли значимым, 45 исследований (47,4%) являются ложноположительными [[35](#_ENREF_35)].

То же самое можно выразить с помощью формулы Байеса.

P(A) – вероятность получить положительный результат исследования, то есть доказательство эффективности вмешательства;

P(H+) – вероятность того, что проверяемая гипотеза истинна (то есть доля истинных гипотез среди проверяемых)

P(H-) – аналогично доля ложных гипотез среди проверяемых

P(A|H+) – чувствительность критерия, статистическая мощность, (1‑β)

P(A|H-) – вероятность ошибки первого рода, α

Формулу Байеса можно записать следующим образом:

$$P\left(A\right)=\frac{P\left(H\_{-}\right)\*P(H\_{-})}{P\left(H\_{-}\right)\*P\left(H\_{-}\right)+P\left(H\_{+}\right)\*P(H\_{+})}$$

Подставив числа из нашего гипотетического примера, мы получим тот же результат.

Можно рассчитать, какова будет вероятность, что обнаруженный положительный результат исследования окажется ложноположительным, для часто применяемых параметров α и β. Результаты приведены в таблице 2.

Таблица 2. Доля ложноположительных результатов из статистически значимых. Значения в таблице приводятся по исследованию [[35](#_ENREF_35)].

| Чувствительность, статистическая мощность (1-β) | Процент значимых результатов являющихся ложноположительными |
| --- | --- |
| P=0,05 | P=0,01 | P=0,001 |
| 80% проверяемых гипотез верно (нулевые гипотезы ложны)  |
| http://www.ncbi.nlm.nih.gov/corehtml/pmc/pmcents/x2003.gif20 | 5,9 | 1,2 | 0,10 |
| http://www.ncbi.nlm.nih.gov/corehtml/pmc/pmcents/x2003.gif50 | 2,4 | 0,5 | 0,05 |
| http://www.ncbi.nlm.nih.gov/corehtml/pmc/pmcents/x2003.gif80 | 1,5 | 0,3 | 0,03 |
| 50% проверяемых гипотез верно (нулевые гипотезы ложны) |
| http://www.ncbi.nlm.nih.gov/corehtml/pmc/pmcents/x2003.gif20 | 20,0 | 4,8 | 0,50 |
| http://www.ncbi.nlm.nih.gov/corehtml/pmc/pmcents/x2003.gif50 | 9,1 | 2,0 | 0,20 |
| http://www.ncbi.nlm.nih.gov/corehtml/pmc/pmcents/x2003.gif80 | 5,9 | 1,2 | 0,10 |
| 10% проверяемых гипотез верно (нулевые гипотезы ложны) |
| http://www.ncbi.nlm.nih.gov/corehtml/pmc/pmcents/x2003.gif20 | 69,2 | 31,0 | 4,30 |
| http://www.ncbi.nlm.nih.gov/corehtml/pmc/pmcents/x2003.gif50 | 47,4 | 15,3 | 1,80 |
| http://www.ncbi.nlm.nih.gov/corehtml/pmc/pmcents/x2003.gif80 | 36,0 | 10,1 | 1,10 |
| 1% проверяемых гипотез верно (нулевые гипотезы ложны) |
| http://www.ncbi.nlm.nih.gov/corehtml/pmc/pmcents/x2003.gif20 | 96,1 | 83,2 | 33,10 |
| http://www.ncbi.nlm.nih.gov/corehtml/pmc/pmcents/x2003.gif50 | 90,8 | 66,4 | 16,50 |
| http://www.ncbi.nlm.nih.gov/corehtml/pmc/pmcents/x2003.gif80 | 86,1 | 55,3 | 11,00 |

Из таблицы видно, что уровень ложноположительных результатов недооценивается исследователями и в большинстве случаев оказывается значительно выше 5%, которые предполагаются при постановке критического уровня p=0,05.

Оценка доли истинно положительных и ложноположительных результатов клинических испытаний, показавших статистически значимый результат, говорящий в пользу эффективности исследуемого лечения, имеет больше значение для принятия решений в здравоохранении. Поскольку положительные результаты клинических испытаний являются основным доказательством эффективности новых препаратов, необходимым для регистрации и допуска на рынок, важно адекватно оценивать силу этих доказательств. Регулирующие органы основываются на уровне p, показанном в исследовании эффективности и рассматривают его как вероятность ложноположительного результата. На самом деле такая вероятность значительно выше. Во многих случаях, особенно, когда доказательством эффективности являются небольшие исследования с малым количеством участников, вероятность ложноположительного результата может даже превышать вероятность истинно положительного результата. Организаторы здравоохранения, рассматривающие подобные доказательства, должны иметь механизм для оценки этих доказательств, и в случае их слабой доказательной силы требовать от производителя более мощных исследований, результат которых с меньшей вероятностью мог бы быть ложноположительным.

Допуск на рынок новых препаратов по ложноположительному результату клинических испытаний может привести к значительным добавочным бессмысленным тратам со стороны пациентов, а также возникновению побочных эффектов в результате применения неэффективного препарата. Кроме того, допуск препарата на рынок часто является ступенью к решению о предоставлении этого препарата пациентам за счет средств общественного здравоохранения. Если такое решение основано на ложноположительном результате, то ограниченные средства здравоохранения будут потрачены впустую. Особенно актуальна эта проблема для орфанных болезней, где испытания новых препаратов проходят на небольшом количестве участников, а стоимость препаратов бывает крайне высокой. Правильная оценка уровня доказательности представленных результатов клинических испытаний является залогом рационального расходования средств.

Целью экспериментальной части данной работы является оценка того, насколько высока вероятность, что полученные в серийных исследованиях положительные результаты могут быть ложноположительными. Для этого предлагается применить записанную выше формулу Байеса к сериям исследований, проведенных перед регистрацией новых лекарственных препаратов. В качестве уровня ошибки первого рода предлагается брать наибольшее значение p, принятого в исследовании в качестве доказательства ложности нулевой гипотезы. Если по каким-то причинам невозможно узнать точное значение полученных p, то рассчитывать, исходя из p=0,05. В качестве уровня ошибки второго рода брать значение 0,2, или, если в исследовании оценивался допустимый уровень β, использовать его. В качестве доли истинных гипотез, среди проверяемых, брать отношение исследований, показавших положительный результат, ко всем исследованиям, проверявшим эффективность данного препарата.

# Глава 2. Насколько часто положительный результат оценки эффективности вмешательства оказывается ложным

## 2.1. Анализ серийных клинических испытаний

Для проведения исследования по оценке вероятности ложноположительного результата в серии исследований в качестве источника были использованы данные с сайта Управления по надзору за качеством пищевых продуктов и медикаментов (Food and Drug Administration, FDA, США) [[50](#_ENREF_50)]. Данный сайт имеет раздел, где представлены отчеты о всех препаратах, прошедших одобрение Управления по надзору за качеством пищевых продуктов и медикаментов [[48](#_ENREF_48)]. Препараты могут проходить процедуру одобрения по разным причинам: новые показания, изменение дозировки препарата, новая форма выпуска, новая технология изготовления, новое сочетание известных активных веществ в одном препарате. Все эти причины предполагают, что первоначальное одобрение по поводу эффективности препарат уже получил ранее, и за прошедшее время могли быть проведены испытания подтверждающие эту эффективность. А препараты, которые не подтвердили свою эффективность в повторных исследованиях, уже были отвергнуты, и не запрашивают одобрения регуляторных органов. Поэтому все препараты, проходящие повторную регистрацию составляют смещенную выборку, в которой доля действительно эффективных препаратов будет выше, чем среди всех испытуемых лекарств.

Чтобы повторные исследования не оказывали влияния, в качестве анализируемых серий исследований были выбраны только те препараты, которые проходят одобрение Управления по надзору за качеством пищевых продуктов и медикаментов как имеющие в своем составе новую молекулярную субстанцию. В таблице на сайте FDA они маркированы единицей. Из таких препаратов выбираются только те, которые согласно предоставленной истории одобрений прошли процедуру одобрения впервые и имеют сопровождающие файлы с информацией, предоставляемой потребителю, сопроводительное письмо об одобрении FDA, обзор и резюме обзора о препарате. В резюме обзора предоставлена информация, какие исследования послужили доказательной базой для подтверждения клинической эффективности препарата.

По указанным названию препарата, названию действующего вещества и синонимам этого названия производится поиск в базе данных регистров клинических испытаний <http://clinicaltrials.gov/> и <https://www.clinicaltrialsregister.eu> для обнаружение всех зарегистрированных исследований данного препарата. Поскольку в исследовании нас интересует оценка эффективности препаратов, то мы не берем в расчет исследования о биодоступности, фармакокинетике и другие виды исследований, а выделяем только контролируемые испытания, в которых на каком-либо этапе проверялась эффективность лечебного вмешательства с использованием данного препарата. Поиск опубликованных результатов испытаний производится в базе данных PubMed [[56](#_ENREF_56)].

Из всех обнаруженных опубликованных результатов испытаний выбираются те испытания, в которых была показана статистически значимая эффективность исследуемого препарата. В этих публикациях производится поиск значений α- и β-ошибок. Если по каким-то причинам обнаружить точные указания не удается, то α-ошибка принимается равной 0,05, а β-ошибка равной 0,2 как наиболее распространенные параметры данных показателей, значения которых считаются приемлемыми [5; 10]. Поскольку во многих исследованиях эффективность препарата измеряется по нескольким исходам, то в качестве значения α-ошибки берется максимальное значение p из всех исходов, которые были расценены исследователями как доказательство эффективности препарата. Доля истинных гипотез среди проверяемых вычисляется как отношение числа испытаний, показавших статистически значимый положительный результат, к общему числу испытаний эффективности, проведенных с данным препаратом. Все эти значения подставляются в формулу Байеса, приведенную в первой главе, и вычисляется вероятность того, что полученный результат анализируемого исследования в данной серии исследований является ложноположительным.

Препарат Zykadia, действующее вещество Ceritinib, синоним LDK378. Был одобрен Управлением по надзору за качеством пищевых продуктов и медикаментов США (FDA) 29 апреля 2014 года по ускоренной процедуре. Препарат предназначен для лечения немелкоклеточного рака легких. Компания-производитель - Novartis Pharmaceuticals Corporation [[61](#_ENREF_61)]. Согласно резюме обзора эффективность и безопасность препарата была доказана в исследовании, которое упоминается под номерами CLDK378X2101 и CLKD378X2101. Официальное название исследования: “A phase I, multicenter, open-label, dose-escalation study of LDK378, administered orally in adult patients with tumors characterized by genetic abnormalities in anaplastic lymphoma kinase (ALK)”. В регистре clinicaltrials.gov это исследование обнаруживается под номером NCT01283516 и заглавием “A Dose Finding Study With Oral LDK378 in Patients With Tumors Characterized by Genetic Abnormalities in Anaplastic Lymphoma Kinase (ALK) ”.

В регистре clinicaltrials.gov при поиске на “Zykadia” и “Ceritinib” не обнаруживается никаких исследований. При поиске на LDK378 обнаруживается 12 записей. Из них одна является опубликованным протоколом исследования и поэтому не принимается во внимание. 8 записей являются исследованиями, в которые продолжается набор участников, следовательно они не являются законченными и не имеют готовых результатов. Три остальных исследования, включая то, основываясь на котором было принято решение о регистрации препарата, не являются контролируемыми и рандомизированными, поэтому не могут быть использованы в расчетах вероятности ложноположительного результата из-за ошибки первого рода. Рандомизированные контролируемые испытания третьей фазы с этим препаратом пока что только начаты, и информация о их регистрации есть в регистре clinicaltrialsregister.eu, но результатов пока что нет.

Препарат Impavido, действующее вещество Miltefosine. Зарегистрирован FDA 19 марта 2014 года как препарат для лечения лейшманиоза. Согласно резюме обзора эффективность препарата подтверждена клиническими испытаниями. Эффективность при лечении висцерального лейшманиоза подтверждена двумя рандомизированными испытаниями, одно из которых было испытанием не меньшей эффективности. Подтверждением эффективности при кожном лейшманиозе являются три испытания, два из которых являлись рандомизированными.

При поиске в регистрах клинических испытаний наиболее полный список, полностью включающий в себя списки других регистров, обнаруживается при поиске на «miltefosine» в регистре clinicaltrials.gov. В списке обнаруживается 24 записи. Из них два исследования находятся на стадии набора участников, в одном исследовании набор еще не начат, и поэтому исключаются из анализа. Среди остальных 21 исследований статус 5 исследований неизвестен, 3 исследования прекращены, 12 исследований завершены, одно указано как активное с завершенной записью участников. Из этих исследований 4 исключено из анализа, поскольку не являются рандомизированными, в одном исследовании сравниваются разные курсы и дозы только исследуемого препарата, еще в трех сравниваются комбинации препаратов с использованием исследуемого, одно исследование не является сравнительным. Таким образом в анализ было включено 12 исследований подходящего дизайна и статуса.

Для пяти [[39](#_ENREF_39); [40](#_ENREF_40); [43](#_ENREF_43); [55](#_ENREF_55); [62](#_ENREF_62)] зарегистрированных исследований опубликованные результаты не обнаруживаются. Для пяти [[21](#_ENREF_21); [25](#_ENREF_25); [28](#_ENREF_28); [34](#_ENREF_34); [37](#_ENREF_37)] исследований опубликованы результаты, но в них не обнаруживается статистически значимых различий между группой с применением мильтефазина и группой сравнения. Два исследования [[22](#_ENREF_22); [23](#_ENREF_23)] показали статистически значимые преимущества в эффективности мильтефазина по сравнению с препаратом сравнения. В одном исследовании [[22](#_ENREF_22)] уровень статистической значимости был равен 0,04, в другом исследовании [[23](#_ENREF_23)] он был равен 0,05. В обоих исследованиях с «положительным» результатом не указан уровень допустимой β-ошибки, поэтому его будем считать равным 0,2. В виду большого влияния публикационного смещения все неопубликованные результаты будем считать не показавшими различий.

Подставим полученные результаты в формулу Байеса из первой главы. Для первого исследования получим:

$P\left(A\right)=\frac{P\left(H\_{-}\right)\*P(H\_{-})}{P\left(H\_{-}\right)\*P\left(H\_{-}\right)+P\left(H\_{+}\right)\*P(H\_{+})}=\frac{0,04\*10/12}{0,04\*\frac{10}{12}+0,8\*\frac{2}{12}}=0,2$

Для второго исследования получим:

$P\left(A\right)=\frac{P\left(H\_{-}\right)\*P(H\_{-})}{P\left(H\_{-}\right)\*P\left(H\_{-}\right)+P\left(H\_{+}\right)\*P(H\_{+})}=\frac{0,05\*10/12}{0,05\*\frac{10}{12}+0,8\*\frac{2}{12}}=0,238$

Таким образом из серии исследований, проведенных для испытания препарата мильтефазин следует, что положительный результат, показанный в исследовании [[22](#_ENREF_22)] с вероятностью 20% является ложноположительным. А положительный результат, показанный в исследовании [[23](#_ENREF_23)] с вероятностью 24% является ложноположительным.

Препарат Osphena, действующее вещество Ospemifene, был одобрен FDA 26 февраля 2013 года для лечения средней или сильной болезненности полового акта у женщин или симптомов вагинальной атрофии во время менопаузы.

При поиске в регистрах клинических испытаний в clinicaltrials.gov обнаруживается 7 записей о клинических испытаниях, в clinicaltrialsregister.eu обнаруживается 3 записи, все 10 записей собраны в регистре ВОЗ.

Из десяти исследований в одном [[33](#_ENREF_33)] изучалась безопасность препарата и не оценивалась его клиническая эффективность, одно исследование [[32](#_ENREF_32)] не являлось контролируемым исследованием. Еще одно исследование [[44](#_ENREF_44)] пока что не начато и не набирает участников. Из семи оставшихся исследований, подходящего дизайна, о двух исследованиях обнаруживается недостаточно информации: нет результатов или отсутствует статистический анализ полученных результатов – поэтому, учитывая возможное публикационное смещение, они будут рассматриваться как не показавшие статистически значимого результата. Об оставшихся пяти исследованиях обнаруживается достаточно информации в отдельных публикациях или в отчете FDA, по которой можно судить о статистической значимости полученных различий между экспериментальными и контрольными группами.

В двух исследованиях по всем оцениваемым исходам был получен результат о различиях в группах, чья статистическая значимость была равна или меньше 0,0001 [[17](#_ENREF_17); [27](#_ENREF_27)]. В одном исследовании максимальное значение p из всех сравниваемых параметров составило p=0,0136 [[13](#_ENREF_13)]. Еще в двух исследования, хотя они рассматриваются как показавшие статистически значимый результат и многие параметры в них показали достаточно низкое значение p, по одному из исходов значение p превышало 0,05 и составило p=0,0853 [[60](#_ENREF_60)] и p=0,080 [[26](#_ENREF_26)]. Ни в одном исследовании не был отдельно указан расчетный уровень β-ошибки, поэтому в расчетах использовано значение 0,2.

Для исследований, показавших значение p<0,0001

$P\left(A\right)=\frac{P\left(H\_{-}\right)\*P(H\_{-})}{P\left(H\_{-}\right)\*P\left(H\_{-}\right)+P\left(H\_{+}\right)\*P(H\_{+})}=\frac{0,0001\*2/7}{0,0001\*\frac{2}{7}+0,8\*\frac{5}{7}}=0,00005$

Для остальных исследований:

$P\left(A\right)=\frac{P\left(H\_{-}\right)\*P(H\_{-})}{P\left(H\_{-}\right)\*P\left(H\_{-}\right)+P\left(H\_{+}\right)\*P(H\_{+})}=\frac{0,0136\*2/7}{0,0136\*\frac{2}{7}+0,8\*\frac{5}{7}}=0,0068$

$P\left(A\right)=\frac{P\left(H\_{-}\right)\*P(H\_{-})}{P\left(H\_{-}\right)\*P\left(H\_{-}\right)+P\left(H\_{+}\right)\*P(H\_{+})}=\frac{0,0853\*2/7}{0,0853\*\frac{2}{7}+0,8\*\frac{5}{7}}=0,0409$

$P\left(A\right)=\frac{P\left(H\_{-}\right)\*P(H\_{-})}{P\left(H\_{-}\right)\*P\left(H\_{-}\right)+P\left(H\_{+}\right)\*P(H\_{+})}=\frac{0,08\*2/7}{0,08\*\frac{2}{7}+0,8\*\frac{5}{7}}=0,0385$

Таким образом для исследований, показавших наименьшее значение p, вероятность ложноположительного результата исследования очень невысока и составляет не более 0,005%. Для исследований с большим значением p в данной серии исследований, вероятность ложноположительного результата составляет 0,68%, 4% и 3,38%, что не превышает традиционный 5% уровень ошибки первого рода.

Препарат Luzu, действующее вещество Luliconazole, одобрен FDA 14 ноября 2013 года. Препарат выпускается в виде 1% крема и предназначен для лечения грибковых поражений стоп, вызванных организмами Trichophyton rubrum и Epidermophyton floccosum.

При поиске клинических исследований в регистре clinicaltrials.gov обнаруживается три записи о зарегистрированных испытаниях, в регистре ВОЗ обнаруживается четыре записи, включающие указанные три, в регистре clinicaltrialsregister.eu не зарегистрировано ни одного испытания данного препарата. Одно [[54](#_ENREF_54)] из четырех исследований не является сравнительным. Одно исследование [[58](#_ENREF_58)] пока еще не завершено и не имеет результатов. Кроме того в резюме обзора упоминаются три исследования эффективности препарата [[59](#_ENREF_59)]. Об одном из этих исследований обнаруживается публикация в базе данных PubMed [[20](#_ENREF_20)], о двух других исследованиях информации обнаружить не удалось. Тот факт, что препарат испытывался в исследованиях, которые не были предварительно зарегистрированы в регистрах клинических испытаний, может говорить об имеющемся публикационном смещении.

Из пяти описанных исследований в трех, хотя и приведены результаты по группам, не указано значение p для статистической значимости различий результатов между препаратом и плацебо. В одном исследовании значение указано как p<0,001 [[20](#_ENREF_20)]. В пятом исследовании [[19](#_ENREF_19)] не указано значение p для сравнения эффективности люликоназола с другими препаратами, проходившими испытание. Таким образом из пяти исследований в 4 статистическая значимость различий неизвестна, и поэтому будет рассчитываться как не значимая.

$P\left(A\right)=\frac{P\left(H\_{-}\right)\*P(H\_{-})}{P\left(H\_{-}\right)\*P\left(H\_{-}\right)+P\left(H\_{+}\right)\*P(H\_{+})}=\frac{0,001\*4/5}{0,001\*\frac{4}{5}+0,8\*\frac{1}{5}}=0,005$

Вероятность ложноположительного результата исследования составляет 0,5%.

Препарат Juxtapid, действующее вещество Lomitapide mesylate, одобрен FDA 21 декабря 2012 года. Препарат предназначен для снижения уровня липопротеинов при различных состояниях в комплексе с другими мерами лечения.

При поиске в регистрах клинических испытаний на “Lomitapide” и его синонимы “AEGR-773” и “BMS-201038” в регистре clinicaltrials.gov обнаруживается 14 исследований, в clinicaltrialsregister.eu находится 3 записи, в регистре ВОЗ выявляется 9 записей, дублирующих два других регистра.

Из 17 исследований 6 еще не завершены и не имеют результатов, в 5 не проводилась оценка эффективности, в двух исследованиях не проводилось сравнения. Четыре исследования являются рандомизироваными контролируемыми испытаниями подходящего дизайна.

В одном исследовании [[46](#_ENREF_46)] статистическая значимость полученных различий результатов в экспериментальных и в контрольной группе составила p<0,001. Другое исследование не показало статистической значимости в сравнении ломитапида с другими группами [[29](#_ENREF_29)]. Для двух исследований представлены полученные результаты [[38](#_ENREF_38); [47](#_ENREF_47)], однако не удалось найти статистической обработки полученных данных и, соответственно, узнать статистическую значимость различий, поэтому она рассматриваются как не показавшие значимых различий.

$P\left(A\right)=\frac{P\left(H\_{-}\right)\*P(H\_{-})}{P\left(H\_{-}\right)\*P\left(H\_{-}\right)+P\left(H\_{+}\right)\*P(H\_{+})}=\frac{0,001\*3/4}{0,001\*\frac{3}{4}+0,8\*\frac{1}{4}}=0,0037$

Для исследования, показавшего в данной серии статистическую значимость, вероятность ложноположительного результата составляет не более 0,37%.

## 2.2. Основные выводы

В данной работе было проанализировано 70 исследований, проводившихся для изучения 5 разных лекарственных препаратов. Вероятность ложноположительного результата в разных исследованиях разных серий получила значения от 0,005% до 24%. Для большинства исследований полученная вероятность ошибки первого рода была в рамках традиционных 5%, такой уровень риска считается приемлемым большинством исследователей. Для одной серии исследований вероятность того, что полученные положительные результаты в некоторых испытаниях из всего ряда являются ложноположительными, превысила 20%. Такой уровень ошибки первого рода для медицинских испытаний является очень высоким.

Полученные результаты показывают, что исследования с достаточно низким значением p (p<0,001) даже в сериях, где они являются единственным исследованием со статистически значимым положительным результатом, показывают небольшую вероятность ошибки первого рода. Исследования, чье значение p было в районе 0,05, который считается приемлемым при анализе одиночного исследования, при анализе в серии показывали значительно большую вероятность ошибки перового рода, которая не всегда может быть приемлемой.

В ходе исследования было выявлено, что меры, принимаемые по проспективной регистрации клинических испытаний пока что недостаточны, и среди опубликованных материалов обнаруживаются результаты исследований, которые не были предварительно зарегистрированы в регистрах клинических испытаний. По всей видимости, некоторые испытания были внесены в регистры ретроспективно, что могло отразить публикационное смещение и касаться только тех исследований, которые показали положительный результат.

Также исследование показало, что не все препараты, прошедшие одобрение FDA, проходили рандомизированые контролируемые испытания. Подобные серии исследований нельзя проверить на вероятность ложноположительного результата, поскольку в самих исследованиях не заложен расчет вероятности ошибки первого рода.

Результаты исследования показывают, что вероятность ложноположительного результата значительно отличается в случае одиночного испытания вмешательства и в случае проведении серии несравнимых между собой испытаний. Для корректной оценки статистической значимости положительного результата, показанного в испытании, нужно учитывать не только данные самого испытания, но и все испытания, проведенные для оценки эффективности данного вмешательства. Из результатов исследования можно сделать предварительный вывод, что традиционный уровень статистической значимости, принятый для медицинских испытаний, равный 0,05 в современных условиях является недостаточным. Современные условия, когда значительно возросло количество проводимых испытаний с каждым исследуемым вмешательством, а также увеличилось количество исходов, по которым оценивается эффективность и безопасность в каждом испытании, требуют снижения допустимого уровня ошибки первого рода в каждом отдельном испытании и в оценке каждого исхода.

Также результаты данного исследования подтверждают общественную важность своевременной публикации результатов всех проведенных клинических испытаний. Только при условии, что результаты всех испытаний будут публиковаться, можно будет корректно оценивать количество гипотез, которые были проверены относительно рассматриваемого вмешательства и оценить долю истинных гипотез среди общего числа.

Оценка вероятности того, что представленный положительный результат испытания эффективности нового вмешательства является ложноположительным, имеет большое экономическое значение для здравоохранения. На доказательствах эффективности вмешательства и оценках размера эффекта базируются расчеты об экономической эффективности расходования ресурсов здравоохранения на одни или другие меры. В случае принятия решения о предоставлении вмешательства за счет средств общественного здравоохранения на основании ложноположительного результата испытания, средства здравоохранения будут потрачены впустую, а экономическая эффективность трат будет нулевой. Оценка того, насколько вероятно, что представленный результат является ложноположительным, может служить основанием для принятия решения о регистрации, закупке, предоставлении рассматриваемого вмешательства. При сравнении нескольких вмешательств между собой, оценка вероятности ложноположительных результатов может использоваться в расчетах эффективности затрат, при этом предполагаемая эффективность затрат будет обратно пропорциональна вероятности ложноположительного результата испытания в серии.

# Глава 3. Практическое применение научно обоснованных методов с учетом возможных смещений

При принятии управленческих и экономических решений организаторы здравоохранения должны руководствоваться доказательствами, основанными на применении строгих научных методов. Научный подход в здравоохранении способствует принятию рациональных решений, в том числе о расходовании ограниченных финансовых ресурсов. Применяя и анализируя результаты современных методов оценки эффективности медицинских вмешательств, включая их в экономический анализ, организаторы здравоохранения должны учитывать ограничения, которые имеют эти методы. Только зная слабые и сильные стороны применяемых методов, можно использовать их правильно.

К слабым сторонам существующих и применяемых научных методов в здравоохранении относятся возможные смещения, которые возникают при оценке эффективности медицинских вмешательств. Организатор здравоохранения при принятии решения о регистрации, закупке или предоставлении какого-либо вмешательства должен помнить, что публикационное смещение способствует увеличению оценки размера эффекта вмешательства. В отдельных публикациях и даже при проведении мета-анализа заявленный размер эффекта вмешательства будет выше, чем в реальности. Это свойство важно учитывать при расчете предполагаемых затрат и при построении предположений о результатах.

Также организаторам здравоохранения необходимо учитывать, что ввиду сложности фиксации побочных эффектов в клинических испытаниях и ограничения их регистрации после начала применения вмешательства, любое вмешательство выглядит более безопасным, чем оно является на самом деле. Всегда существует риск отдаленных последствий, который может быть выяснен со значительной задержкой во времени (до десятков лет). Кроме того существует риск того, что вмешательство имеет побочный эффект, незначительный и нефиксируемый в масштабах клинических исследований, но масштаб которого будет иметь значение при массовом применении вмешательства. При планировании потенциальной выгоды от предоставления вмешательства, всегда нужно учитывать, что расходы, связанные с побочными эффектами, будут больше, чем это можно предположить по результатам клинических испытаний.

Зачастую решения, которые принимают организаторы здравоохранения, напрямую затрагивают интересы провайдеров тех или иных медицинских услуг. Применение лишних вмешательств, увеличение числа и стоимости вмешательств, введение в оборот новых сложных и дорогостоящих вмешательств всегда будет в интересах тех, кто получает деньги за выполнение вмешательств. Заинтересованной стороной могут быть производители оборудования, фармацевтические компании, медицинские учреждения или врачебные организации. В интересах системы здравоохранения тратить ограниченные финансовые ресурсы только на эффективные вмешательства и избегать бессмысленных и лишних трат. Принимая решение о предоставлении вмешательства за счет средств общественного здравоохранения следует учитывать, что заинтересованная сторона будет преувеличивать эффект от дополнительных новых вмешательств по сравнению с уже применяемыми. Организатор здравоохранения должен уметь отделять важные плоды прогресса в медицине от навязанных провайдерами лишних вмешательств.

Все корректные научные доказательства эффективности медицинских вмешательств в том или ином виде опираются на статистику, так как испытания проводятся на выборках, а выводы делаются относительно популяции. Организатор здравоохранения, принимающий решения, должен понимать весомость различных статистических доказательств. Понимание, насколько часто среди испытаний вмешательств встречаются ложноположительные результаты, может помочь избежать решений о лишних бессмысленных затратах и значительно сэкономить средства общественного здравоохранения.

Частично уменьшить влияние смещений при применении научных методов в здравоохранении можно с помощью изменения законодательства. Поскольку пациенты, участвующие в клинических испытаниях, подвергаются определенному риску, а от результатов испытания зависят здоровье многих людей, общество в праве требовать, чтобы исследователи не утаивали информацию о проведенных испытаниях. Требование проспективной регистрации всех клинических испытаний в базе данных, доступной для просмотра населением, содержится в Хельсинкской декларации, которая является этическим руководством для проведения экспериментов на людях [[4](#_ENREF_4)]. На этом основании можно рекомендовать внесение подобного требования в федеральные законы. Вполне вероятно, что введение чувствительных штрафов для юридических лиц, организующих или сотрудничающих в проведении незарегистрированных испытаний на людях, будет достаточно, чтобы значительно повысить уровень клинических испытаний, зарегистрированных в открытых регистрах до набора первого участника.

Также актуально законодательно обязать исследователей своевременно публиковать результаты испытаний. Поскольку одна только проспективная регистрация испытаний не спасает от публикационного смещения в том числе в результате более поздней публикации отчетов об испытаниях, показавших отрицательный результат. Оговоренные в законе разумные сроки на публикацию всех результатов проведенных исследований позволят с большей уверенностью говорить о том, что при принятии решений учтены все имеющиеся на определенный момент времени данные об эффективности и безопасности анализируемого вмешательства.

Этап одобрения и регистрации препаратов перед их выходом на рынок является одной из мер, которую предпринимают, чтобы не допустить на рынок заведомо бессмысленные и неэффективные препараты. Второй важный этап, при котором требуется оценка эффективности, это принятие решения о предоставлении препарата за счет средств общественного здравоохранения. В обоих случаях основным доказательством эффективности препарата является статистически значимое различие между экспериментальной и контрольной группами в клиническом испытании. Но, как было показано в данной работе, уровень статистической значимости одного исследования не говорит о реальной вероятности ошибки первого рода, то есть ложноположительного результата испытания. Только при учете всех испытаний, проведенных с использованием анализируемого препарата, можно сделать правильный вывод о том, насколько вероятно, что представленный положительный результат испытания является истинным.

При принятии решения о регистрации или предоставлении нового препарата организаторам здравоохранения следует запрашивать информацию о всех проведенных испытаниях, в том числе не относящихся непосредственно к тому назначению, по которому препарат проходит регистрацию или предлагается к закупке. Проанализировав все завершенные испытания, узнав среди них долю положительных и отрицательных результатов, включив в анализ заложенный при проведении исследования допустимый уровень β-ошибки и полученное значение p, по методике, предложенной в данной работе, можно вычислить уровень ошибки первого рода для конкретного препарата в конкретной серии исследований. Принимая решение, следует руководствоваться именно таким вычисленным уровнем ошибки первого рода для серии, а не значением p в отдельно взятом исследовании эффективности. При регистрации или закупке препарата стоит воздерживаться от принятия решения на основании лишь одного испытания, показавшего эффективность. Анализ практики FDA по одобрению лекарственных препаратов показал, что решение о регистрации препарата часто принимается на основании минимального количества испытаний – медианное значение составляет 2 исследования (интерквартильный интервал 1-2,5). А 36,8% показаний для применения препарата были одобрены на основании лишь одного испытания [[63](#_ENREF_63)]. Такое качество доказательств вызывает настороженность у специалистов, а также может вводить в заблуждение врачей и пациентов, которые рассматривают сам факт одобрения препарата FDA как подтверждение того, что препарат имеет убедительные научные доказательства эффективности.

Исследователям и организаторам здравоохранения следует рассматривать вероятность ошибки первого рода не как некоторое стандартное пороговое значение, одинаковое для всех, а только в контексте оценки конкретного препарата или вмешательства. Допустимый уровень ошибок первого и второго рода зависит от цены, которую система здравоохранения, общество и конкретный пациент будут платить в случае принятия ложноположительного результата или отвержении истинно положительного результата. Так в случае появления лекарства от опасной болезни, которая на данный момент не имеет никаких средств лечения, допустим достаточно высокий уровень ошибки первого рода, потому что ситуация, в которой из-за сомнений в истинности положительного результата будет упущен эффективный способ лечения, для пациента и всего общества будет значительно хуже, чем если будет предпринято неэффективное вмешательства. Особенно это верно в тех случаях, когда предлагаемое новое вмешательство не является опасным, вредным, болезненным или очень дорогостоящим.

В противоположном случае, когда болезнь не представляет большой опасности и уже имеет проверенный метод лечения, принятие нового метода должно происходить с большей строгостью. В такой ситуации переход на неэффективный препарат, показавший ложноположительный результат, будет нести больше рисков – отказ от прежнего эффективного лечения, отсутствие пользы, лишние затраты – чем отказ от препарата с несколько большей эффективностью, если положительный результат был истинным. Для таких случаев можно снижать допустимый порог α-ошибки и требовать от производителя более весомых и точных доказательств эффективности вмешательства. Для общества будет полезнее, если производители будут проводить меньшее количество исследований, но с большей численностью участников и с большей точностью.

Неправильно думать, что оценка эффективности вмешательств интересна только на этапе регистрации препарата в масштабе национального фармацевтического рынка. На каждом уровне организации здравоохранения – на уровне государства, региона, и даже на уровне отдельного медицинского учреждения – правильная оценка эффективности вмешательства, с учетом имеющихся смещений, может дать положительный экономический результат. Ведь решение о том, какой препарат или вмешательство следует закупать и предоставлять, по своему принимается на всех уровнях организации здравоохранения. На государственному уровне – это решения о регистрации препарата, о включении его в программу госгарантий, включение в списки и перечни, предоставляемые населению или отдельным группам населения на территории всей страны. На региональном уровне правильная оценка эффективности вмешательств дает возможность рационально принимать решения о включении вмешательства в региональные программы, о проведении региональных профилактических и скрининговых мероприятий, о выборе при закупках препаратов и оборудования. На уровне отдельного медицинского учреждения оценка эффективности вмешательств может иметь влияние на решения о выборе закупаемых препаратов и оборудования, рациональную организацию процесса лечения, стратегию развития учреждения.

Преодоление существующих ограничений в применении научных методов при оценке эффективности медицинских вмешательств имеет больше значение и на международном уровне. Существует целый ряд международных инициатив, направленных на уменьшение влияния смещений при оценке эффективности.

Одной из важнейших международных организаций в сфере оценки эффективности медицинских вмешательств является Кокрановское сотрудничество [[64](#_ENREF_64)]. В организацию входит более 28000 ученых из 120 стран. Цель организации – способствовать принятию в области медицины и здравоохранения информированных решений, основанных на доказательствах. Кокрановское сотрудничество создает систематические обзоры и мета-анализы, которые публикуются в их базе данных. Поскольку мета-анализ нескольких рандомизированных контролируемых испытаний обладает наибольшим уровнем доказательности в медицине, он меньше подвержен существующим смещениям.

Другой важной международной инициативой в области клинических испытаний медицинских вмешательств является Хельсинкская декларация [[4](#_ENREF_4)]. В ней сформулированы важные положения об этических обязательствах исследователей и редакторов перед обществом и пациентами. Отдельно указано, что все испытания должны быть зарегистрированы и доступны населению для ознакомления. А также обязательство исследователей обнародовать результаты испытаний с участием людей, вне зависимости от полученного результата: положительного, отрицательного или спорного. Декларация призывает не принимать к публикации отчеты об исследованиях, которые противоречат принципам декларации. Аналогичную меру по отношению к исследователям, не регистрирующим испытания в регистрах, применяет Международный комитет редакторов медицинских журналов (ICMJE). Распространяемые комитетом требования к рукописям указывают: «Комитет требует, и рекомендует всем редакторам медицинских журналов требовать, регистрации всех клинических испытаний в открытых регистрах клинических испытаний до или во время включения в исследование первого пациента» (перевод с английского) [[57](#_ENREF_57)].

В январе 2013 года началась международная кампания под названием AllTrials, целью которой является регистрация всех клинических испытаний и публикация всех их результатов. В ходе кампании был опубликован подробный план, как можно добиться регистрации всех клинических испытаний, публикации всех их результатов и контроля за выполнением этих правил. Для достижения этой цели потребуется объединение усилий многих сторон: регистраторов, регулирующих органов, организаторов и спонсоров клинических испытаний, университетов и институтов, редакций научных журналов, профессиональных сообществ, исследователей, пациентов. Поскольку регистрация и публикация всех результатов клинических испытаний очень выгодна для всего общества, можно надеяться, что кампания добьется поставленных целей.

# Заключение

Развитие медицины и здравоохранения во всем мире определяет все большую востребованность строгих научных доказательств в процессе принятия решений. Научные доказательства являются той рациональной основой, на которой принимаются решения о распределении ограниченных общественных ресурсов в системе здравоохранения.

Процесс принятия экономических и организационных решений в сфере здравоохранения, основанных на научных доказательствах эффективности медицинских вмешательств, несовершенен и имеет некоторые слабые места. В данной работе были рассмотрены имеющиеся несовершенства, проанализированы существующие пути решения и предложены рекомендации для повышения надежности и обоснованности принимаемых решений.

В первой главе данной работы были проанализированы смещения, оказывающие влияние на оценку эффективности медицинских вмешательств, и по каждому смещению предложены возможные пути решения существующей проблемы. В анализ вошли публикационное смещение, смещение при фиксации побочных эффектов вмешательств, тенденция к рекомендации лишних вмешательств, проблема недостаточной унификации медицинских вмешательств, проблема принятия ложноположительного результата испытания эффективности вмешательства. В работе описан алгоритм вычисления вероятности ложноположительного результата испытания в серии испытаний одного и того же вмешательства.

Во второй главе работы представлено исследование семидесяти клинических испытаний, проведенных для оценки эффективности пяти лекарственных препаратов. Представлены расчеты о вероятности ложноположительного результата среди представленных исследований. Сделаны выводы о влиянии возможных ложноположительных результатов на экономические и организационные решения в здравоохранении. Даны рекомендации по оценке и преодолению возможного смещения в сторону принятия ложноположительных результатов клинических испытаний.

В третьей главе данной работы рассматривается, какие возможности имеет организатор здравоохранения при использовании результатов применения научных методов, с учетом имеющихся смещений. Описаны возможные пути уменьшения влияния смещений на принятие решений в здравоохранении. Даны рекомендации по изменению законодательной базы для уменьшения смещений в сфере клинических испытаний. Показана важность правильного использования результатов применения научных методов организаторами здравоохранения на разных уровнях – государственном, региональном, учрежденческом. Также приведены примеры международных инициатив, направленных на повышение уровня доказательности и достоверности научных методов, применяемых в здравоохранении и медицине.

Проведенное исследование показывает, насколько важно оценивать вероятность ложноположительного результата испытания по эффективности вмешательства. При проведении серии испытаний, вероятность ложноположительного результата зависит от совокупности всех гипотез об эффективности вмешательства, проверяемых во всех исследованиях. Результаты исследования показывают, что в современных условиях при оценке эффективности медицинского вмешательства необоснованно судить о вероятности ложноположительного результата исследования по одному лишь значению p. Необходимо также учитывать долю положительных результатов испытаний в серии и заложенную в исследовании вероятность ошибки второго рода. Вероятность ложноположительного результата в одном испытании из серии может значительно отличаться и, что важно для принятия решений, может превышать традиционно считающимся допустимым уровень в 5%.

Повышение обоснованности и экономической эффективности принимаемых решений является важной задачей для организаторов здравоохранения всех уровней. Умение правильно анализировать и учитывать возможные смещения в оценке эффективности медицинских вмешательств является важным навыком для всех, кто связан с принятием решения о регистрации новых медицинских вмешательств или о предоставлении их пациентам за счет средств общественного здравоохранения.

Организаторам здравоохранения важно не только осознавать наличие проблемы подобных смещений, но и прикладывать усилия в сторону изменения ситуации и повышения обоснованности всех принимаемых решений. Повышение точности в оценке эффективности новых медицинских вмешательств приведет к повышению рациональности и результативности в расходовании средств здравоохранения.

Для повышения достоверности и точности оценки эффективности медицинских вмешательств необходимо:

1. Добиться проспективной регистрации всех клинических испытаний и последующей публикации всех результатов. Для достижения цели возможно внесение изменений в законодательство.

2. При принятии решений о регистрации и предоставлении новых вмешательств всегда учитывать завышение размера эффекта вмешательства и занижение количества побочных эффектов.

3. При оценке эффективности вмешательств и проведении экономических расчетов, основанных на эффективности вмешательств, рассчитывать вероятность ложноположительного результата исследования исходя из всех проведенных испытаний данного вмешательства.

4. Повышать точность проводимых клинических испытаний, часто за счет увеличения численности участников.

5. Поддерживать международные инициативы, направленные на повышение обоснованности и доказательности при принятии решений в медицине и здравоохранении.

Итогом данных мероприятий должны стать правильные и обоснованные решения в сфере здравоохранения, принятие которых повысит эффективность использования средств.

# Литература

1. Все клинические испытания зарегистрированы. Все их результаты доступны. Перевод на русский язык манифеста кампании AllTrials. URL: <http://www.alltrials.net/wp-content/uploads/2014/02/AllTrials-Manifesto-Russian-version.pdf> (дата обращения: 12.05.2014).

2. Реестр выданных разрешений на клинические испытания медицинских изделий. URL: <http://www.roszdravnadzor.ru/national_foreign_medprod/Reg_med_products/register_for_clinic> (дата обращения: 12.05.2014).

3. Реестр выданных разрешений на проведение клинических исследований лекарственных препаратов [РКИ]. URL: <http://grls.rosminzdrav.ru/CIPermitionReg.aspx> (дата обращения: 12.05.2014).

4. Хельсинская декларация Всемирной медицинской ассоциации, 1964. URL: <http://www.psychepravo.ru/law/int/helsinkskaya-deklaraciya.htm> (дата обращения: 12.05.2014).

5. Власов В.В. Эпидемиология. — Москва ‌: Издательский дом "ГЭОТАР-МЕД", 2004. — 462 с.

6. Гланц С. Медико-биологическая статистика. Пер. с англ. — Москва ‌: Практика, 1999. — 459 с.

7. Горецкая М.Р. Доказательная медицина: новые подходы к образованию и практике здравоохранения // Бизнес медицина. 2005. Т. 7.

8. Гринхальх Т. Основы доказательной медицины. Пер. с англ. — Москва ‌: ГЭОТАР-Медиа, 2006. — 240 с.

9. Корнышева Е.А., Платонов Д.Ю., Родионов А.А., Шабашов А.Е. Эпидемиология и статистика как инструменты доказательной медицины. Издание второе исправленное и дополненное — Тверь. , 2009. — 80 с.

10. Реброва О.Ю. Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTICA. — Москва ‌: Медиа Сфера, 2002. — 312 с.

11. Флетчер Р., Флетчер С., Вагнер Э. Клиническая эпидемиология. Основы доказательной медицины. Пер. с англ. — Москва ‌: Медиа Сфера, 1998. — 352 с.

12. Хачатрян Г.Р., Лунева А.В. Инновации в медицине и определение их ценности в зарубежных странах: обзор литературы // Медицинские технологии. Оценка и выбор. 2013. Т. 4. № 14. — C. 23-29.

13. Bachmann G.A., Komi J.O. Ospemifene effectively treats vulvovaginal atrophy in postmenopausal women: results from a pivotal phase 3 study // Menopause. 2010. Т. 17. № 3. — C. 480-6.

14. Biller-Andorno N., Jüni P. Abolishing Mammography Screening Programs? A View from the Swiss Medical Board // The New England Journal of Medicine. 2014.

15. Egger M., Smith G.D., Schneider M., Minder C. Bias in meta-analysis detected by a simple, graphical test // BMJ. 1997. Т. 315. № 7109. — C. 629-634.

16. Esserman L.J., Thompson I.M., Jr, Reid B. Overdiagnosis and overtreatment in cancer: An opportunity for improvement // JAMA. 2013. Т. 310. № 8. — C. 797-798.

17. Goldstein S.R., Bachmann G.A., Koninckx P.R., Lin V.H., Portman D.J., Ylikorkala O. Ospemifene 12-month safety and efficacy in postmenopausal women with vulvar and vaginal atrophy // Climacteric. 2014. Т. 17. № 2. — C. 173-82.

18. Ioannidis J.P. Why most published research findings are false // PLoS Med. Т. 2. № 8.

19. Jerajani H., Janaki C., Kumar S., Phiske M. Comparative assessment of the efficacy and safety of sertaconazole (2%) cream versus terbinafine cream (1%) versus luliconazole (1%) cream in patients with dermatophytoses: a pilot study // Indian J Dermatol. 2013. Т. 58. № 1. — C. 34-8.

20. Jones T.M., Jarratt M.T., Mendez-Moguel I., Paz N., Grekin S.K., Cognata Smith C., Kaur M. A randomized, multicenter, double-blind, vehicle-controlled study evaluating the efficacy and safety of luliconazole cream 1% once daily for 7 days in patients aged >/= 12 years with tinea cruris // J Drugs Dermatol. 2014. Т. 13. № 1. — C. 32-8.

21. Lopez L., Cruz C., Godoy G., Robledo S.M., Velez I.D. Thermotherapy effective and safer than miltefosine in the treatment of cutaneous leishmaniasis in Colombia // Rev Inst Med Trop Sao Paulo. 2013. Т. 55. № 3.

22. Machado P.R., Ampuero J., Guimaraes L.H., Villasboas L., Rocha A.T., Schriefer A., Sousa R.S., Talhari A., Penna G., Carvalho E.M. Miltefosine in the treatment of cutaneous leishmaniasis caused by Leishmania braziliensis in Brazil: a randomized and controlled trial // PLoS Negl Trop Dis. 2010. Т. 4. № 12. — C. e912.

23. Magerl M., Rother M., Bieber T., Biedermann T., Brasch J., Dominicus R., Hunzelmann N., Jakob T., Mahler V., Popp G., Schakel K., Schlingensiepen R., Schmitt J., Siebenhaar F., Simon J.C., Staubach P., Wedi B., Weidner C., Maurer M. Randomized, double-blind, placebo-controlled study of safety and efficacy of miltefosine in antihistamine-resistant chronic spontaneous urticaria // J Eur Acad Dermatol Venereol. 2013. Т. 27. № 3. — C. e363-9.

24. McGauran N., Wieseler B., Kreis J., Schuler Y.B., Kolsch H., Kaiser T. Reporting bias in medical research - a narrative review // Trials. 2010. Т. 11. — C. 37.

25. Mohebali M., Fotouhi A., Hooshmand B., Zarei Z., Akhoundi B., Rahnema A., Razaghian A.R., Kabir M.J., Nadim A. Comparison of miltefosine and meglumine antimoniate for the treatment of zoonotic cutaneous leishmaniasis (ZCL) by a randomized clinical trial in Iran // Acta Trop. 2007. Т. 103. № 1. — C. 33-40.

26. Portman D., Palacios S., Nappi R.E., Mueck A.O. Ospemifene, a non-oestrogen selective oestrogen receptor modulator for the treatment of vaginal dryness associated with postmenopausal vulvar and vaginal atrophy: A randomised, placebo-controlled, phase III trial // Maturitas. 2014.

27. Portman D.J., Bachmann G.A., Simon J.A. Ospemifene, a novel selective estrogen receptor modulator for treating dyspareunia associated with postmenopausal vulvar and vaginal atrophy // Menopause. 2013. Т. 20. № 6. — C. 623-30.

28. Rubiano L.C., Miranda M.C., Muvdi Arenas S., Montero L.M., Rodriguez-Barraquer I., Garcerant D., Prager M., Osorio L., Rojas M.X., Perez M., Nicholls R.S., Gore Saravia N. Noninferiority of miltefosine versus meglumine antimoniate for cutaneous leishmaniasis in children // J Infect Dis. 2012. Т. 205. № 4. — C. 684-92.

29. Samaha F.F., McKenney J., Bloedon L.T., Sasiela W.J., Rader D.J. Inhibition of microsomal triglyceride transfer protein alone or with ezetimibe in patients with moderate hypercholesterolemia // Nat Clin Pract Cardiovasc Med. 2008. Т. 5. № 8. — C. 497-505.

30. Schroter S., Glasziou P., Heneghan C. Quality of descriptions of treatments: a review of published randomised controlled trials // BMJ Open. 2012. Т. 2. № 6.

31. Shamliyan T.A., Kane R.L. Availability of results from clinical research: failing policy efforts // J Epidemiol Glob Health. 2014. Т. 4. № 1. — C. 1-12.

32. Simon J., Portman D., Mabey R.G., Jr. Long-term safety of ospemifene (52-week extension) in the treatment of vulvar and vaginal atrophy in hysterectomized postmenopausal women // Maturitas. 2014. Т. 77. № 3. — C. 274-81.

33. Simon J.A., Lin V.H., Radovich C., Bachmann G.A. One-year long-term safety extension study of ospemifene for the treatment of vulvar and vaginal atrophy in postmenopausal women with a uterus // Menopause. 2013. Т. 20. № 4. — C. 418-27.

34. Soto J., Rea J., Balderrama M., Toledo J., Soto P., Valda L., Berman J.D. Efficacy of miltefosine for Bolivian cutaneous leishmaniasis // Am J Trop Med Hyg. 2008. Т. 78. № 2. — C. 210-1.

35. Sterne J.A., Davey Smith G. Sifting the evidence-what's wrong with significance tests? // BMJ. 2001. Т. 322. № 7280. — C. 226-31.

36. Sterne J.A., Sutton A.J., Ioannidis J.P., Terrin N., Jones D.R., Lau J., Carpenter J., Rucker G., Harbord R.M., Schmid C.H., Tetzlaff J., Deeks J.J., Peters J., Macaskill P., Schwarzer G., Duval S., Altman D.G., Moher D., Higgins J.P. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials // BMJ. 2011. Т. 343. — C. d4002.

37. Sundar S., Sinha P.K., Rai M., Verma D.K., Nawin K., Alam S., Chakravarty J., Vaillant M., Verma N., Pandey K., Kumari P., Lal C.S., Arora R., Sharma B., Ellis S., Strub-Wourgaft N., Balasegaram M., Olliaro P., Das P., Modabber F. Comparison of short-course multidrug treatment with standard therapy for visceral leishmaniasis in India: an open-label, non-inferiority, randomised controlled trial // Lancet. 2011. Т. 377. № 9764. — C. 477-86.

38. AEGR-733 and Atorvastatin 20 mg vs. Monotherapy in Moderate Hypercholesterolemia. URL: <http://clinicaltrials.gov/ct2/show/results/NCT00474240> (дата обращения: 12.05.2014).

39. Clinical Trial of Miltefosine to Treat Mucosal Leishmaniasis. URL: <http://clinicaltrials.gov/ct2/show/NCT01377974> (дата обращения: 12.05.2014).

40. Clinical Trial to Assess the Safety and Efficacy of Sodium Stibogluconate (SSG) and AmBisome® Combination, Miltefosine and AmBisome® and Miltefosine Alone for the Treatment Visceral Leishmaniasis in Eastern Africa. URL: <http://clinicaltrials.gov/ct2/show/NCT01067443> (дата обращения: 12.05.2014).

41. ClinicalTrials.gov. A service of the U.S. National Institutes of Health. URL: <http://clinicaltrials.gov/ct2/home> (дата обращения: 12.05.2014).

42. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. Available from <http://www.cochrane-handbook.org.Higgins> JPT G.S. ‌: The Cochrane Collaboration, 2011.

43. Combination Therapy in Indian Visceral Leishmaniasis. URL: <http://clinicaltrials.gov/ct2/show/NCT00523965> (дата обращения: 12.05.2014).

44. Effects of Ospemifene on Pelvic Vascularity and Blood Flow. URL: <http://clinicaltrials.gov/ct2/show/NCT02010580> (дата обращения: 12.05.2014).

45. EU Clinical Trials Register. URL: https://<http://www.clinicaltrialsregister.eu/ctr-search/search> (дата обращения: 12.05.2014).

46. Evaluate Low Doses of AEGR-733 on Hepatic Fat Accumulation by MRS. URL: <http://clinicaltrials.gov/ct2/show/results/NCT00559962> (дата обращения: 12.05.2014).

47. Evaluate Safety and Efficacy of AEGR-733 and Atorvastatin vs Atorvastatin Monotherapy in Hypercholesterolemia. URL: <http://clinicaltrials.gov/ct2/show/results/NCT00690443> (дата обращения: 12.05.2014).

48. FDA Approved Drug Products. Drug Approcal Reports. URL: <http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm?fuseaction=Reports.ReportsMenu> (дата обращения: 12.05.2014).

49. Food and Drug Administration Modernization Act of 1997. URL: <http://www.gpo.gov/fdsys/pkg/PLAW-105publ115/pdf/PLAW-105publ115.pdf>.

50. Food and Drug Administration, FDA (Управление по надзору за качеством пищевых продуктов и медикаментов ). URL: <http://www.fda.gov/>.

51. Institute for Quality and Efficiency in Health Care, IQWIG (Институт качества и эффективности в здравоохранении в Германии). URL: https://<http://www.iqwig.de/en/home.2724.html>.

52. International Clinical Trials Registry Platform Search Portal. URL: <http://apps.who.int/trialsearch/> (дата обращения: 12.05.2014).

53. National Institute for Health and Care Excellence, NICE (Национальный институт здоровья и клинического совершенствования Великобритании). URL: <http://nice.org.uk/>.

54. Open-Label Pharmacokinetics (PK)/Safety Study of Luliconazole Solution, 10% in Distal Subungual Onychomycosis. URL: <http://clinicaltrials.gov/ct2/show/study/NCT01044381> (дата обращения: 12.05.2014).

55. Phase III, Study of Three Short Course Combination Regimens (Ambisome®, Miltefosine, Paromomycin) Compared With AmBisome® Alone for the Treatment of Visceral Leishmaniasis in Bangladesh. URL: <http://clinicaltrials.gov/ct2/show/NCT01122771> (дата обращения: 12.05.2014).

56. PubMed database. URL: <http://www.ncbi.nlm.nih.gov/pubmed> (дата обращения: 12.05.2014).

57. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals (ICMJE Recommendations), 2013. URL: <http://www.icmje.org/icmje-recommendations.pdf> (дата обращения: 12.05.2014).

58. Safety and Efficacy of Luliconazole Solution, 10% in Subjects With Mild to Moderate Onychomycosis (SOLUTION). URL: <http://clinicaltrials.gov/ct2/show/study/NCT01431820> (дата обращения: 12.05.2014).

59. Summary Review for Luliconazole. Application number: 204153Orig1s000. URL: <http://www.accessdata.fda.gov/drugsatfda_docs/nda/2013/204153Orig1s000SumR.pdf> (дата обращения: 12.05.2014).

60. Summary Review for Osphena/ospemifene. Application number: 203505Orig1s000. URL: <http://www.accessdata.fda.gov/drugsatfda_docs/nda/2013/203505Orig1s000SumR.pdf> (дата обращения: 12.05.2014).

61. Summary Review for Zykadia (ceritinib) Capsules. Application number: 205755Orig1s000. URL: <http://www.accessdata.fda.gov/drugsatfda_docs/nda/2014/205755Orig1s000SumR.pdf> (дата обращения: 12.05.2014).

62. Treatment of Cutaneous Leishmaniasis With a Combination of Miltefosine and Antimony. URL: <http://clinicaltrials.gov/ct2/show/NCT01380301> (дата обращения: 12.05.2014).

63. Downing N.S., Aminawung J.A., Shah N.D., Krumholz H.M., Ross J.S. Clinical trial evidence supporting FDA approval of novel therapeutic agents, 2005-2012 // JAMA. 2014. Т. 311. № 4. — C. 368-77.

64. The Cochrane Collaboration. URL: <http://www.cochrane.org/> (дата обращения: 12.05.2014).