

Тема 3. Введение в доказательную медицину. Оценка диагностического теста.

Доказательная медицина (ДМ) – это технология сбора, анализа, обобщения и интерпретации медицинской информации, позволяющая принимать научно доказательные решения по профилактике, диагностике, лечению заболеваний и организации здравоохранения.

В настоящее время доказательная медицина является основополагающим инструментом для принятия решения о выборе медицинской технологии медицинских работников. Центром принятия решений является не мнение авторитета или укоренившиеся традиции, а специалист-медик – ответственный и компетентный, информированный и критически мыслящий.

История медицины учит, что выбор клинического решения, основанный только на теоретических данных и на результатах неконтролируемых экспериментальных исследований, ненадежен. В связи с этим решение врача по возможности должно опираться на результаты контролируемых исследований. Такие исследования, если они хорошо спланированы и проведены на высоком методологическом уровне, сводят к минимуму вероятность возникновения ошибок, характерных для неконтролируемых исследований.

В основе доказательной медицины лежит эмпирический подход, поскольку главный ее принцип состоит в непосредственном наблюдении за изменением состояния больных на фоне того или иного вмешательства.

Любая наука базируется на «научно обоснованных фактах» и доказательствах. Исследователь может сказать, что имеются убедительные основания считать какую-либо теорию верной, если ее положения подтверждаются результатами экспериментальных исследований. Однако в доказательной медицине данный термин имеет совершенно иной смысл, характеризуя лишь серию наблюдений. С позиции доказательной медицины утверждение о том, что имеются убедительные основания считать тот или иной метод лечения эффективным, обычно означает, что частота излечения среди пациентов, получавших его, превышает таковую у не получавших. При этом различие должно превышать определенную пороговую величину. Для того чтобы убедиться в этом, проводят статистический анализ.

Доказательная медицина с ее методологией проведения рандомизированных клинических, диагностических, эпидемиологических исследований и метаанализа имеет большое значение для клинической практики, однако, следует помнить, что любой полученный результат всегда относится лишь к той или иной группе больных, участвовавших в исследовании.

Клиническое исследование — научное исследование с участием людей, которое проводится с целью оценки эффективности и безопасности нового лекарственного препарата или расширения показаний к применению уже известного лекарственного препарата.

Клинические исследования во всем мире являются неотъемлемым этапом разработки препаратов, который предшествует его регистрации и широкому медицинскому применению. В ходе клинических исследований новый препарат изучается для получения данных о его эффективности и безопасности. На основании этих данных уполномоченный орган здравоохранения принимает решение о регистрации препарата или отказе в регистрации.

Концепции доказательной медицины распространяются по трем основным направлениям:

- 1) *Разработка клинических рекомендаций, описывающих действия специалистов-медиков в определенной клинической ситуации.* При принятии клинического решения специалисты должны опираться на эти рекомендации с учетом индивидуальных особенностей больного. Составителями рекомендаций выступают профессиональные врачебные ассоциации или правительственные организации, образующие экспертные группы. При внедрении в практику клинические рекомендации могут обеспечить, помимо

всего прочего, колоссальный экономический эффект прежде всего за счет отказа от дорогостоящих малоинформативных диагностических методов и лекарственных средств, не обладающих доказательной клинической ценностью.

2) *Издание специализированных обучающих и справочных бумажных и электронных журналов, руководств, книг и Интернет-ресурсов.* Первым таким изданием стал журнал «ASP Journal Club» Американской коллегии врачей (1991). С 1996 г. стало выходить дочернее издание – журнал «Evidence-Based Medicine». В середине 90-х годов прошлого столетия в России начал издаваться «Международный журнал медицинской практики», который стал первым серьезным источником информации по доказательной медицине на русском языке. Наиболее детальное и строгое руководство по доказательной медицине в мире — «Кокрановское руководство по подготовке систематических обзоров».

3) *Формирование базы данных систематических обзоров рандомизированных контролируемых исследований.* Данное направление развивается в рамках Кокрановского сообщества (Cochrane Collaboration). Это международное содружество ученых, выявляющих, систематизирующих и обобщающих результаты всех когда-либо опубликованных рандомизированных контролируемых исследований. С помощью метаанализа ученые суммируют данные, полученные в ходе различных исследований по одной проблеме. В результате такого синтеза информации удастся объективно, на основе статистических выкладок, оценить степень полезности различных лечебных, диагностических и профилактических вмешательств. Кокрановская электронная библиотека находится на сайте в Интернете [1], доступна русская версия. Систематические обзоры публикуются в ежегодном справочнике Clinical Evidence, русский перевод которого – «Доказательная медицина. Ежегодный справочник.» Также на русском языке в Интернете представлены электронные ресурсы доказательной медицины издательства «Медиа Сфера» [2].

В 1996 г. опубликован важнейший международный документ – CONSORT, описывающий правила подготовки публикаций результатов рандомизированных контролируемых исследований. Эти правила приняты в качестве стандарта научных публикаций ведущими международными медицинскими журналами и профессиональными организациями медицинских редакторов. Суть правил состоит в том, чтобы дать возможность читателю судить о научной достоверности статьи и надежности сделанных выводов.

Статьи попадают в Кокрановскую базу данных только при условии, что они заявлены в тексте как рандомизированные и (или) контролируемые.

Контролируемое клиническое исследование, в котором исследуемое лекарственное средство, эффективность и безопасность которого до конца еще не изучены, сравнивают с препаратом, эффективность и безопасность которого хорошо известны (препарат сравнения). Это может быть плацебо, стандартная терапия или отсутствие лечения вообще. В **неконтролируемом** (несравнительном) исследовании группа контроля / сравнения (группа испытуемых, принимающих препарат сравнения) не используется. В более широком смысле под контролируемым исследованием имеется в виду всякое исследование, в котором контролируются (по возможности минимизируются или исключаются) потенциальные источники систематических ошибок (т. е. оно проводится в строгом соответствии с протоколом, мониторируется и т.д.).

Степень доказательности данных исследования зависит от его дизайна (типа).

Для получения надежных научно обоснованных результатов необходимо правильное планирование структуры исследования (дизайн) и адекватный статистический анализ.

Планирование исследования состоит из 2 этапов: определение типа исследований и определение объема выборок.

Классификация медицинских исследований может быть проведена по:

Цели исследования:

– выдвигающее гипотезу

– проверяющее гипотезу

времени проведения исследования:

– одномоментное (поперечное) однократное обследование участников или объектов наблюдения;

– динамическое (продольное) многократное обследование участников или объектов наблюдения;

Наличию или отсутствию вмешательства – активное (использование метода лечения) и пассивное (наблюдение за естественным течением заболевания);

Соотношению времени сбора данных и формированию выборок:

– проспективное (группы формируются до сбора данных)

– ретроспективное (группы формируются после сбора данных).

В зависимости от **количества исследовательских центров**, в которых проводится исследование в соответствии с единым протоколом, исследования бывают одноцентровыми и многоцентровыми. Если исследование проводится в нескольких странах, его называют международным.

В результате комбинации представленных выше характеристик исследования формируются следующие виды структуры исследования (в порядке возрастания доказательности):

1. Описание отдельных случаев;
2. Описание серии случаев;
3. Исследование случай-контроль;
4. Аналитическое одномоментное исследование;
5. Проспективное когортное (популяционное) исследование;
6. Рандомизируемое контролируемое испытание медицинских вмешательств (методов лечения, профилактики);
7. Мета-анализ – обобщение результатов нескольких рандомизированных клинических испытаний

Описание отдельных случаев – описание редкого наблюдения, «классического случая» или нового феномена. В описании случая (синонимы: случай из практики, история заболевания, описание единичного случая) ведется исследование лечения и исхода у одного человека.

В **исследовании серии случаев** наблюдают несколько индивидуумов, обычно получающих одинаковое лечение, без использования контрольной группы.

В **исследовании случай-контроль** (синоним: исследование сходных случаев) сравнивают людей с определенным заболеванием или исходами («случай») с людьми из этой же популяции, не страдающими данным заболеванием, или у которых не наблюдался данный исход («контроль»), с целью выявления связи между исходом и предшествующему воздействию определенных риск-факторов.

Проспективное исследование проводится с делением участников на группы, которые будут или не будут получать исследуемое лекарственное средство, до того, как наступили исходы. В отличие от него, в ретроспективном (историческом) исследовании изучаются исходы проведенных ранее клинических исследований, т.е. исходы наступают до того, как начато исследование.

В **параллельном исследовании** сравниваются две или более группы испытуемых, одна или более из которых получают исследуемый препарат, а одна группа является контрольной. В некоторых параллельных исследованиях сравнивают различные виды лечения, без включения контрольной группы. (Такой дизайн называют дизайном независимых групп).

Когортное исследование – это обсервационное исследование, в котором выделенную группу людей (когорту) наблюдают в течение некоторого времени. Исходы у испытуемых в разных подгруппах данной когорты, тех кто подвергся или не подвергся

(или подвергался в разной степени) лечению исследуемым препаратом сравниваются. В проспективном когортном исследовании когорты составляют в настоящем и наблюдают их в будущем. В ретроспективном (или историческом) когортном исследовании когорту подбирают по архивным записям и прослеживают их исходы с того момента по настоящее время.

Различают уровни доказательности исследования (от 3 до 5-7), их обозначают латинскими цифрами, чем меньше цифра, тем большую достоверность имеют данные исследования.

1. Категория I - к ней относятся хорошо разработанные, крупные, рандомизированные, контролируемые исследования, данные мета-анализа или систематических обзоров.

2. К категории II относятся когортные исследования и исследования типа случай-контроль.

3. К категории III принято относить неконтролируемые исследования и консенсусы специалистов.

Практические рекомендации по диагностике и лечению опираются как на данные исследований, так и на экстраполяции этих данных. С учетом этого обстоятельства рекомендации подразделяются на 3-5 уровней, которые принято обозначать латинскими буквами - A, B, C, D, E.

Рекомендации уровня A базируются на результатах исследований, относимых к I категории доказательности и, следовательно, отличаются наиболее высоким уровнем достоверности. Достоверность рекомендаций уровня B также довольно высока - при их формулировании используются материалы исследований II категории или экстраполяции исследований I категории доказательности. Рекомендации уровня C строятся на основании неконтролируемых исследований и консенсусов специалистов (III категории доказательности) или содержат экстраполяции рекомендаций I и II категории.

Уровень	Терапия / Профилактика, Этиология/Риск
1a	Систематические Обзоры Рандомизированных Клинических Испытаний РКИ
1b	Отдельные РКИ
2a	Систематические Обзоры Когортных Исследований
2b	Отдельные Когортные испытания
2c	Экологические исследования
3a	Систематические Обзоры исследований Случай-контроль
3b	Отдельные исследования Случай-контроль
4	Серии случаев
5	Мнение экспертов без точной критической оценки, или основанный на физиологии

Основные определения математической статистики, встречающиеся в области клинических исследований

Прежде чем перейти к математической интерпретации задач, возникающих в области клинических исследований, приведем некоторые общие определения, встречающиеся в области планирования и анализа результатов клинических исследований.

Статистика – наука, изучающая методы сбора и интерпретации числовых данных. Главная цель статистики — получение осмысленных заключений из несогласованных (подверженных разбросу) данных. Так как индивидуумам присуща врожденная изменчивость признаков (например, рост, масса тела и т.п.) и, кроме того, биологические признаки могут случайным образом меняться во времени, природа клинических данных, характеризующихся разбросом или вариацией, диктует необходимость их статистического оценивания. Еще одна причина, по которой применение статистических методов к данным клинических исследований становится необходимым, – случайные ошибки измерения клинических показателей. На языке математики величина любого варьирующего признака является *переменной случайной величиной*, а ее конкретные значения принято называть *вариантами*.

Важной задачей при проведении статистического анализа клинических данных является определение одного или нескольких признаков, которые в данном клиническом исследовании будут адекватно оценивать сравниваемый эффект. Вообще словом «эффект» независимо от его медицинского смысла мы будем называть любые проявления действия изучаемого препарата (или метода лечения), которые выбраны исследователем для демонстрации его эффективности, безопасности и т.д. Показатели определенного таким образом эффекта обладают межиндивидуальной вариабельностью.

Статистику еще часто определяют как науку принятия разумных решений перед лицом неопределенности. При этом двум категориям задач статистики уделяют особое внимание: статистическое оценивание и проверка статистических гипотез. Первая задача делится на точечное и интервальное оценивание параметров распределения. Вообще статистические задачи появляются тогда, когда необходимо дать наилучшие, в некотором смысле, ответы по ограниченному числу наблюдений. Если бы число наблюдений не было ограниченным, можно было точно определить параметры распределений и сравнить их, при этом никакой статистической задачи не было бы. Если в ходе исследований мы могли бы изучить все объекты интересующей нас совокупности (например, всех больных с определенным заболеванием), то можно было бы сказать, что мы имеем дело со сплошным изучением *генеральной совокупности*. На самом деле обследовать все объекты совокупности удастся редко, обычно приходится изучать лишь *выборку*, надеясь, что эта выборка достаточно хорошо отражает свойства изучаемой совокупности. При этом также возникают важные статистические задачи: *случайный отбор вариант* из генеральной совокупности и *представительность* выборки, а также определение *необходимого объема выборки* для формирования статистически значимого заключения по результатам проведенных исследований.

Все статистические методы исходят из предположения, что данные извлечены из совокупности случайно. А это значит, что вероятность оказаться выбранным для всех членов совокупности должна быть одинакова. Случайным должно быть и отнесение пациента к той или иной сравниваемой группе, т.е. каждый пациент должен иметь равный шанс попасть в любую группу в исследовании. Предназначенные для решения этой статистической задачи методы называются *методами рандомизации*. Рандомизация не только уравнивает вероятность получения пациентом различных сравниваемых воздействий, но и позволяет формировать группы, сходные с точки зрения прогностических факторов. Таким образом, рандомизация обеспечивает подбор больных так, чтобы контрольная группа ни в чем не отличалась от экспериментальной, кроме изучаемого метода лечения. Но и этого оказывается недостаточно. Тесно связана с проблемой рандомизации и так называемая проблема слепоты исследования. Для того чтобы ни врач,

ни исследователь, ни пациент не могли каким-либо образом влиять на получаемые результаты, используется такое понятие, как *слепота* исследования. Например, если позволяют клинические особенности данного исследования, часто применяется так называемый *двойной слепой метод*, когда ни врач, ни пациент не знают, какой из методов лечения был применен.

Важным вопросом является и вопрос *представительности* (или *репрезентативности*) выборки по отношению ко всей популяции, из которой она отбиралась. Обычно, если выборка извлечена из совокупности случайным образом и имеет достаточно большой объем, средние характеристики пациентов в выборке практически такие же, как в соответствующей популяции. На практике большинство групп пациентов, включенных в различные клинические исследования, представляют собой смещенные выборки. Это связано с особенностями включения пациентов в исследование: часто пациенты включаются потому, что находятся на лечении в центре, проводящем исследование, или потому, что, с точки зрения исследователя, представляют собой интересный клинический случай. В принципе такое отсутствие репрезентативности не приводит к каким-то неправильным выводам. Однако исследователь должен четко понимать, на какую популяцию реально могут быть распространены результаты, полученные в таком исследовании.

Планирование медицинских экспериментов

При планировании исследований очень важно сформулировать его *цель*. Если целью проводимого исследования является установление различий (или преимущества) методов лечения, математически данный вопрос решается обычно с помощью проверки *статистического критерия (или теста)*. Применяемые для этого процедуры связаны с формулированием статистических гипотез. Иногда для решения этой задачи применяют и метод *доверительных интервалов*.

Особенностью планирования клинических исследований является то, что исследователь никогда не имеет в своем распоряжении всей популяции (генеральной совокупности) для проведения исследования и обычно имеет дело только с выборкой из этой совокупности. При этом особую важность приобретают задачи планирования, например, определение объема выборки, которого оказалось бы достаточно для формирования статистически значимого заключения о различиях (или отсутствии таких различий) в эффекте по результатам проведенного исследования. На практике часто на этапе планирования исследования эта задача не решается строго, а планируемый объем назначается на основе прошлого опыта проведения аналогичных исследований. При этом в процессе статистического анализа может оказаться, что полученных данных или недостаточно для статистически достоверного ответа на вопросы, ради которых проводилось исследование, или их структура не соответствует цели исследования. Исследователям надо иметь в виду, что неэтичными являются исследования, имеющие чрезмерно большую численность включенных пациентов, так и исследования слишком малого объема.

Чтобы ответить на вопросы, поставленные перед клиническим исследованием, с помощью различных критериев приходится проверять нулевую гипотезу об отсутствии эффекта, сравнивая при этом выборочные средние, доли, кривые выживаемости и т.д. Вывод об отсутствии таких различий тесно связан с понятием чувствительность критерия. **Чувствительностью** критерия называют его способность обнаружить различия. Чтобы оценить чувствительность критерия, нужно задать величину различий, которые он должен выявлять. Если в результате проверки гипотезы о существовании различий был сделан вывод об их отсутствии, необходимо проверить, была ли чувствительность критерия достаточной для обнаружения таких различий. Чувствительность зависит не только от величины различий, но и от разброса данных и объема выборки. Если посмотреть на объем выборки, чем он больше, тем меньшие различия окажутся статистически значимыми. Таким

образом, появляется возможность заранее оценивать численность выборок, необходимых для выявления эффекта.

Анализ и сравнение диагностических тестов

Диагностический показатель представляет собой биологический феномен, возникающий или изменяющийся при патологии, который может быть обнаружен с помощью стандартизованного диагностического метода.

Для того чтобы избежать путаницы в методических особенностях выявления диагностических показателей, в лабораторной медицине используется еще одно понятие – **диагностический тест**. Диагностический тест представляет собой определение диагностического показателя с помощью конкретного лабораторного метода, аналитические параметры которого остаются на постоянном уровне при условии контроля качества его выполнения.

«Золотой» стандарт - такой диагностический тест, который максимально точно в настоящее время определяет наличие или отсутствие определенной болезни у пациента.

Результатом использования диагностического теста является получение данных. Вне зависимости от характера получаемых данных (качественных или количественных) результаты каждого теста могут быть условно разделены на положительные и отрицательные. Обычно в качестве такой условной границы используется одна из границ нормы, однако могут быть выделены другие границы. Хотя большинство тестов могут использоваться для решения нескольких клинических задач, аналитические характеристики лабораторного теста обычно исследуются именно при постановке диагноза. Процесс установления диагноза несовершен. Мы можем лишь предполагать, что диагноз верен, нежели утверждать это со всей определенностью. Раньше врачи выражали степень уверенности в клиническом диагнозе предваряя его словами «исключено» или «возможно». Современные клиницисты все чаще выражают свою уверенность в диагнозе через вероятности. Следовательно, врач должен понимать математическую суть диагностической ценности тестов в различных ситуациях. Как правило это помогает врачу уменьшить степень неопределенности того или иного диагностической ситуации. В ряде случаев – точнее оценить эту неопределенность, а иногда лишь осознать степень своей неуверенности в диагнозе.

При оценке аналитических параметров теста он используется на двух группах обследованных: больных с заболеванием и контрольной популяции. Соотношение между результатами тестирования и верным диагнозом представлено в таблице 1.

Таблица 3.1.

Диагностический тест	Наличие болезни	
	ЕСТЬ	НЕТ
ПОЛОЖИТЕЛЬНЫЙ	ИП	ЛП
ОТРИЦАТЕЛЬНЫЙ	ЛО	ИО

ИП– истинно положительный, ИО - истинно отрицательный, ЛО – ложно-отрицательный, ЛО - ложно-положительный

Всего возможны четыре варианта толкования результатов теста – два истинных и два ложных. Наиболее важными параметрами любого теста являются аналитическая точность, чувствительность и специфичность.

Таблица 3.2.

Точность (диагностическая эффективность)	=	ИП+ИО
		ИП+ИО+ЛП+ЛО
Чувствительность (Ч)	=	ИП
		ИП +ЛО
Специфичность (С)	=	ИО
		ИО+ЛП
Фактор риска положительного результата (отношение правдоподобия +)	=	Ч
		(1-С)
Фактор риска отрицательного результата (отношение правдоподобия –)	=	(1-Ч)
		С
Предсказание положительного результата	=	(1-С)
		ИО+ЛП
Предсказание отрицательного результата	=	ИО
		ИО+ЛП

Для сравнения разных диагностических тестов между собой используется понятие **аналитической точности**, представляющей собой число истинных результатов к числу всех результатов обследования. С помощью этого показателя можно сопоставить различные методы определения одного показателя при обследовании одной популяции больных.

Чувствительность представляет собой число корректных положительных результатов обследования к общему числу обследованных больных. Вопрос на который отвечает данная характеристика диагностического теста: насколько хорош тест для выявления пациентов, имеющих данное состояние?

Специфичность указывает на число корректных отрицательных результатов обследования к общему числу контрольной группы. Чувствительный тест следует выбрать, если есть риск пропустить заболевание при неясной диагностической картине или сузить рамки диагностического поиска, исключив ряд частых причин с помощью высокочувствительных тестов. В последнем случае для врача чувствительный тест наиболее информативен, когда он дает отрицательный результат.

Специфичные тесты нужны для подтверждения диагноза, предложенного на основании других данных. Вопрос на который отвечает данная специфичность диагностического теста: насколько хорош тест для правильного исключения пациентов, не имеющих данное состояние?

Для того чтобы уточнить вероятность постановки диагноза с помощью теста в стандартной клинической ситуации используется понятие **предсказательной ценности** (predictive value) – если у человека тест положительный, какова вероятность того, что у него действительно есть данное заболевание?

Предсказание положительного результата теста (ППР) определяется как вероятность того, что пациент болен, если получен положительный результат теста. ППР непосредственно зависит от специфичности теста. При высоких значениях этого показателя можно считать, что полученные результаты подтверждают предполагаемый диагноз.

Предсказание отрицательного результата (ПОР) определяет вероятность того, что обследованный здоров, если получен отрицательный результат теста. Чем чувствительнее тест, тем выше его ПОР. То есть, возрастает уверенность врача в том, что отрицательные результаты теста отвергают наличие заболевания. Важно подчеркнуть, что ППР и ПОР являются вторичными показателями, так как прямо зависят от соотношения числа больных в группе с заболеванием и числа лиц контрольной группы. Чем больше было обследовано больных по сравнению с контрольной группой, тем выше ППР, и наоборот, чем меньше больных было обследовано для определения аналитических параметров теста, тем выше ПОР.

Отношение правдоподобия указывает на значимость теста для повышения уверенности относительно положительного диагноза, т.е. насколько более вероятно то, что тест будет положительным у человека с заболеванием по сравнению со здоровым? Высокое отношение правдоподобия указывает на высокую ценность теста, но из этого не обязательно следует то, что положительный результат теста является хорошим критерием наличия болезни.