КАЗНУ ИМ. АЛЬ-ФАРАБИ · ФАКУЛЬТЕТ МЕДИЦИНЫ И ЗДРАВООХРАНЕНИЯ

Доказательная медицина · Лекция 9

Оценка диагностических тестов в доказательной медицине

Чувствительность, специфичность, прогностическая ценность, отношение правдоподобия и ROC-анализ

Раушанова А.М.
и.о. профессора кафедры ЭБиДМ ФМиЗ КазНУ им. аль-Фараби

Алматы · 2026

О ЗАНЯТИИ

Цель и вопросы занятия

ЦЕЛЬ ЗАНЯТИЯ

Научить оценивать точность диагностических тестов, рассчитывать и интерпретировать показатели их информативности и применять результаты диагностических исследовани конкретному пациенту.

РАССМАТРИВАЕМЫЕ ВОПРОСЫ

1. Диагностический вопрос по PICO: индексный и эталонный тест
2. Таблица 2×2: чувствительность и специфичность
3. Прогностическая ценность и роль распространённости заболевания
4. Отношения правдоподобия; претестов посттестовая вероятность
5. ROC-кривая и выбор порога отсечения
6. Дизайн и систематические ошибки диагностических исследований; QUAD STARD
7. Диагностические и скрининговые тест

ОЖИДАЕМЫЕ РЕЗУЛЬТАТЫ

Что ожидаем по итогам занятия

Знать

- показатели точности теста и формулы их расчёта
- зависимость прогностической ценности от распространённости
- требования к дизайну диагностического исследования и типичные ошибки

Уметь

- строить таблицу 2×2 и рассчитывать Se, Sp, PPV, NPV, LR
- переводить претестовую вероятность в посттестовую
- читать ROC-кривую и сравнивать тесты по AUC

Владеть

- навыком критической оценки диагностическо исследования (QUADAS объяснения результата пациенту

Результаты занятия — часть итоговых компетенций цикла: сформулировать вопрос, найти и критически оценить доказательств применить их на практике.

РАЗМИНКА

Положительный тест — значит, болен?

Тест выявляет заболевание у **90%** больных и даёт отрицательный результат у **80%** здоровых
Заболевание подтверждается у **1 из 10** пациентов, направленных на обследование. У вашего
пациента тест положительный.

Какова вероятность, что пациент действительно болен?

А	Б	В	Г
≈ 90%	≈ 80%	≈ 50%	≈ 33%

Голосуем руками. Ответ получим через несколько слайдов.

ШАГ 1 · ВОПРОС

Диагностический вопрос по PICO

P
ПАЦИЕНТЫ
Взрослые с подозрением на тромбоз глубоких вен

I
ИНДЕКСНЫЙ ТЕСТ
Изучаемый тест: D-димер

C
ЭТАЛОННЫЙ ТЕСТ
Компрессионное УЗИ вен

O
ИСХОД
Точность: Se, Sp, P
NPV, LR

У взрослых с подозрением на ТГВ насколько точно D-димер позволяет исключить тромбоз по сравнению с компрессионным УЗИ?

Эталонный тест («золотой стандарт») — луч
доступный способ установить диагноз:
гистология, ангиография, катамнестическое
наблюдение. Точность индексного теста все
оценивают относительно него.

ОСНОВА РАСЧЁТОВ

Таблица 2 × 2

	БОЛЕЗНЬ ЕСТЬ	БОЛЕЗНИ НЕТ
Тест +	**a** истинно положительные	**b** ложноположительные
Тест –	**c** ложноотрицательные	**d** истинно отрицательные

Как читать

Столбцы — результат **эталонного** теста, строки — результат **индексного** теста.

Все больные: **a + c**. Все здоровые: **b + d**.

Зелёные ячейки — верные результаты, оранжевые — ошибки теста.

Все показатели точности рассчитывают из эти четырёх чисел.

СВОЙСТВА ТЕСТА

Чувствительность и специфичность

Чувствительность (Se)

$$Se = a / (a + c)$$

Доля больных, у которых тест положительный.

SnNout: отрицательный результат высокочувствительного теста помогает **исключить** болезнь.

Специфичность (Sp)

$$Sp = d / (b + d)$$

Доля здоровых, у которых тест отрицательн

SpPin: положительный результат высокоспецифичного теста помогает **подтвердить** болезнь.

Se и Sp — свойства самого теста: они мало зависят от распространённости болезни, но зависят о **спектра пациентов** и **порога отсечения**.

ВОПРОС ВРАЧА И ПАЦИЕНТА

Прогностическая ценность результата

Положительная (PPV)

$$PPV = a / (a + b)$$

Вероятность болезни при **положительном** результате теста.

Отрицательная (NPV)

$$NPV = d / (c + d)$$

Вероятность отсутствия болезни при **отрицательном** результате.

Пациент спрашивает не «какова чувствительность теста?», а «я болен?». На этот вопрос отвечает прогностическая ценность, но она **зависит от распространённости** заболевания в той группе, где применяют тест.

РАСЧЁТ НА ПРИМЕРЕ

Ответ на разминку

	БОЛЕЗНЬ ЕСТЬ	БОЛЕЗНИ НЕТ	ВСЕГО
Тест +	**90** истинно положительные	**180** ложноположительные	**270** a + b
Тест −	**10** ложноотрицательные	**720** истинно отрицательные	**730** c + d
Всего	**100**	**900**	**1000**

1 000 пациентов, распространённость 10%, Se 90%, Sp 80%

90%

Se = 90 / 100

80%

Sp = 720 / 900

33%

PPV = 90 / 270

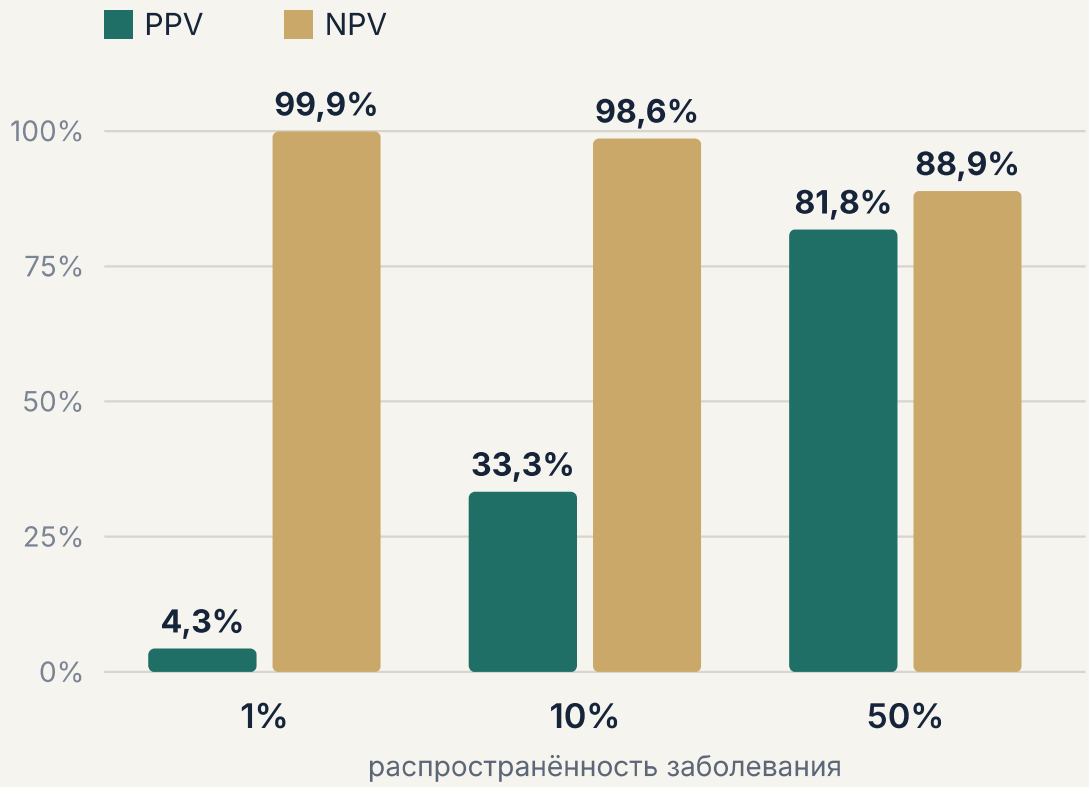
98,6%

NPV = 720 / 73

Правильный ответ — **Г, ≈ 33%**. Два из т положительных результатов ложные: 18 здоровых против 90 больных.

РОЛЬ РАСПРОСТРАНЁННОСТИ

Один тест — разная прогностическая ценнос



Тот же тест: Se 90%, Sp 80%

При скрининге, где болезнь редка, большинство положительных результатов ложные: PPV около 4%.

В специализированном стационаре, где бо встречается часто, PPV выше, но падает N

Вывод: PPV и NPV из статьи нельзя перенс на свою популяцию без пересчёта.

ПЕРЕНОС НА ПАЦИЕНТА

Отношения правдоподобия (LR)

Положительное

$$LR+ = Se / (1 - Sp)$$

Пример: 0,90 / 0,20 = **4,5**

Отрицательное

$$LR- = (1 - Se) / Sp$$

Пример: 0,10 / 0,80 = **0,125**

LR+	LR−	Влияние на вероятность
> 10	< 0,1	Большое, часто решающее
5–10	0,1–0,2	Умеренное
2–5	0,2–0,5	Небольшое
1–2	0,5–1	Минимальное
1	1	Тест ничего не меняет

LR показывает, во сколько раз результат теста чаще встречается у больных, чем у здоровых. Как и Se, Sp зависит от распространённости.

ПЕРЕНОС НА ПАЦИЕНТА

От претестовой к посттестовой вероятности

1 Претестовая вероятность — из клинической картины или распространённости

→

2 Шансы до теста: p / (1 − p)

→

3 Шансы после теста: шансы до × LR

→

4 Вероятность посл теста: шансы / (1 + шанс

Тест положительный

10% → шансы 0,11 × 4,5 = 0,5 → **33%**

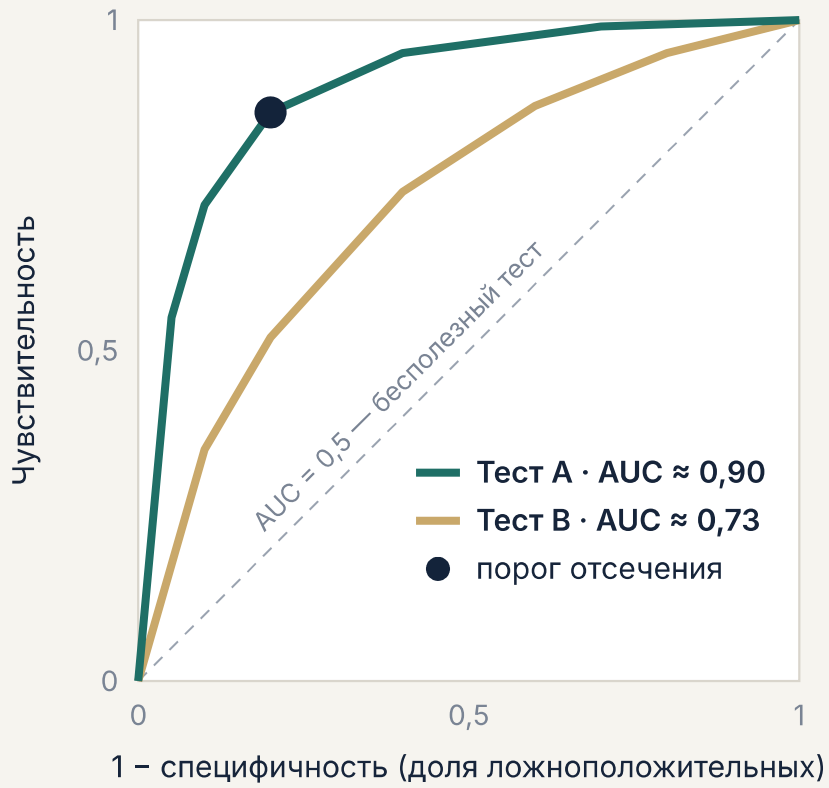
Тест отрицательный

10% → шансы 0,11 × 0,125 = 0,014 → **1,4%**

Тот же ответ, что и по таблице 2 × 2, но без пересчёта таблицы для каждого пациента. На практик удобна **номограмма Фагана** или онлайн-калькулятор.

КОЛИЧЕСТВЕННЫЕ ТЕСТЫ

ROC-кривая и порог отсечения



Порог отсечения

Снижаем порог → чувствительность растёт, специфичность падает. Повышаем → наоборот.

Выбор порога зависит от цены ошибки: что опаснее пропустить болезнь или ложно её диагностировать?

Площадь под кривой (AUC)	Информативность
0,9–1,0	Отличная
0,8–0,9	Хорошая
0,7–0,8	Удовлетворительная
0,5–0,7	Низкая

ДИЗАЙН

Как устроено диагностическое исследование

1. Последовательные пациенты с **подозрением** на заболевание
2. → Индексный тест
3. → Эталонный тест — **всем** участникам
4. → Слепая интерпретация результатов
5. → Таблица 2 × расчёт показателе 95% ДИ

Оптимальный дизайн

Поперечное исследование: оба теста выполняют одновременно у пациентов, которым тест действительно нужен в практике.

Диагностическое «случай–контроль»

Сравнивают явно больных со здоровыми добровольцами. Самые трудные для диагно пациенты выпадают, и точность теста **завышается**.

КРИТИЧЕСКАЯ ОЦЕНКА

Систематические ошибки и QUADAS-2

Ошибка	Суть
Спектра	Участники не похожи на пациентов из реальной практики
Верификации	Эталонный тест получают не все, а в основном с положительным индексным
Включения	Результат индексного теста входит в эталонный диагноз
Неослеплённой оценки	Интерпретатор знает результат другого теста

QUADAS-2: четыре домена

1. Отбор пациентов
2. Индексный тест
3. Эталонный тест
4. Поток пациентов и время

Стандарт отчёта — STARD 2015. Не путать: S
— полнота описания, QUADAS-2 — риск оши

ПРИМЕНЕНИЕ

Диагностика и скрининг

Диагностический тест

У пациента есть жалобы или симптомы.

Цель — подтвердить или исключить конкретный диагноз.

Распространённость в группе выше, PPV выше.

Скрининговый тест

Человек считает себя здоровым.

Цель — раннее выявление в популяции.

Сначала высокочувствительный тест, затем подтверждение высокоспецифичным.

Скрининг оправдан, если ранее лечение улучшает исходы (критерии Вильсона–Юнгнера, ВОЗ, 196
Эффект скрининга искажают **ошибка опережения**, **ошибка длительности** и **гипердиагностика**, поэтому его оценивают в РКИ по смертности.

ПРАКТИКУМ

Работа в группах: оцениваем тест

Условие

400 пациентов с подозрением на заболевание. Эталонный тест подтвердил его у 100.

Новый тест оказался положительным у **85** больных и у **60** здоровых.

Формат: группы по 3–4 человека, 15 минут, затем разбор.

1. Постройте таблицу 2 × 2.
2. Рассчитайте Se, Sp, PPV и NPV.
3. Рассчитайте LR+ и LR−. Насколько информати тест?
4. Как изменится PPV, если тест применить там, распространённость 5%?
5. Объясните положительный результат «пациент одну минуту.

ИТОГИ

Ключевые выводы

01 Точность теста оценивают относительно эталонного теста по таблице 2 × 2.

02 Чувствительный тест помогает исключить болезнь (SnNout), специфичный — подтвердит (SpPin).

03 PPV и NPV зависят от распространённости: при редкой болезни большинство положител результатов ложные.

04 Отношения правдоподобия переводят претестовую вероятность в посттестовую для конкретного пациента.

05 Достоверность исследования оценивают по QUADAS-2: опасны ошибки спектра, верификации и неослеплённой оценки.

ИСТОЧНИКИ

Рекомендуемая литература

Флетчер Р., Флетчер С., Вагнер Э. *Клиническая эпидемиология. Основы доказательной медицины.* — М.: Меди
Сфера, 1998. — Глава «Диагностика».

Straus S.E., Glasziou P., Richardson W.S., Haynes R.B. *Evidence-Based Medicine: How to Practice and Teach EBM.* 5
— Elsevier, 2018.

Whiting P.F. et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Me*
2011;155(8):529–536.

Bossuyt P.M. et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ.*
2015;351:h5527.

ОБСУЖДЕНИЕ

Вопросы для обсуждения материала

1. Почему высокочувствительный тест хорош для исключения диагноза, а высокоспецифичный — для его подтверждения?
2. Почему прогностическая ценность одного и того же теста различается в поликлинике и в специализированном стационаре?
3. Se и Sp получены на тяжёлых госпитализированных больных и здоровых добровольцах. Можно ли доверять этим цифрам?
4. Как выбрать порог отсечения для тест опасное, но излечимое заболевание?
5. Почему отношения правдоподобия уд переносить на своих пациентов, чем P NPV?
6. Всегда ли ранняя диагностика полезна Обсудите гипердиагностику на пример скрининга.
7. Сформулируйте по PICO диагностичес вопрос из вашей практики и выберите эталонный тест.

ДЛЯ ПРЕПОДАВАТЕЛЯ

Ключ к практикуму

	БОЛЕЗНЬ ЕСТЬ	БОЛЕЗНИ НЕТ	ВСЕГО
Тест +	**85** истинно положительные	**60** ложноположительные	**145** a + b
Тест −	**15** ложноотрицательные	**240** истинно отрицательные	**255** c + d
Всего	**100**	**300**	**400**

При распространённости 5% PPV снижается до **18,3%**, NPV растёт до **99,0%**. Тест умеренно информативен: лучше исключает, чем подтверждает диагноз.

85%
Se = 85 / 100

80%
Sp = 240 / 300

58,6%
PPV = 85 / 145

94,1%
NPV = 240 / 255

4,25
LR+ = 0,85 / 0,20

0,19
LR− = 0,15 / 0,80