

# «Подводные камни» статистического анализа и клинической интерпретации полученных оценок на примере пациентов с хронической болезнью почек

## Часть III: Оценка информативности биомаркеров

А.Б. Зулькарнаев<sup>1\*</sup>, Е.В. Паршина<sup>2</sup>

<sup>1</sup> ГБУЗ МО "Московский областной научно-исследовательский клинический институт (МОНКИ) им. М.Ф. Владимирского", хирургическое отделение трансплантологии и диализа, 129110, Москва, ул. Щепкина, 61/2, корпус 6, Российская Федерация

<sup>2</sup> Санкт-Петербургский государственный университет, Клиника высоких медицинских технологий им. Н.И. Пирогова, отделение амбулаторного диализа, 198103, Санкт-Петербург, набережная реки Фонтанки, 154, Российская Федерация

## Pitfalls of statistical analysis and clinical interpretation of the estimates on the example of patients with chronic kidney disease

### Part III: Evaluating the informativeness of biomarkers

A.B. Zulkarnaev<sup>1\*</sup>, E.V. Parshina<sup>2</sup>

<sup>1</sup> M.F. Vladimirovsky Moscow Regional Research and Clinical Institute ("MONIKI"), Surgical Department of Transplantology and dialysis, 61/2 Schepkina str., build. 6, Moscow, 129110, Russian Federation

<sup>2</sup> Saint Petersburg State University Hospital, Department of dialysis outpatients, 154 Fontanka emb., Saint-Petersburg, 198103, Russian Federation

**Ключевые слова:** биомаркер, чувствительность, специфичность, прогностическая ценность положительного результата, прогностическая ценность отрицательного результата, распространенность заболевания, статистика

#### Реферат

В эпоху современной медицины исследователи перешли от практики простого эмпирического описания частоты встречаемости симптомов при различных заболеваниях к представлению определенных оценок. Практикующие доктора нередко ошибочно интерпретируют показатели информативности биомаркеров, приводимые в научных публикациях. Наиболее распространенными оценками являются чувствительность ( $Se$ ) – вероятность наличия маркера при условии наличия заболевания, специфичность ( $Sp$ ) – вероятность отсутствия маркера при условии отсутствия заболевания, прогностическая ценность положительного результата теста ( $PPV$ ) – вероятность наличия болезни при условии наличия маркера, прогностическая ценность отрицательного результата ( $NPV$ ) – вероятность отсутствия заболевания при условии отсутствия маркера.

Означает ли, что большое значение  $Se$ ,  $Sp$ ,  $PPV$  или  $NPV$ , а также выявленная статистическая значимость сопряженности маркера и исхода будет свидетельствовать, что маркер эффективен? Далеко не всегда. Статистическая значимость связи маркера с исходом – лишь необходимое, но не доста-

Адрес для переписки: Зулькарнаев Алексей Батыргараевич  
e-mail: 7059899@gmail.com

Corresponding author: Alexey B. Zulkarnaev  
e-mail: 7059899@gmail.com

\* ORCID: <https://orcid.org/0000-0001-5405-7887>

точное условие. На особенности практического применения маркера непосредственное влияние оказывают частота его встречаемости ( $P_M$ ) и распространенность заболевания ( $P_D$ ) в анализируемой популяции. Тот факт, что некоторые условные показатели информативности могут принимать большие значения даже в условиях отсутствия реальной сопряженности маркера и исхода (определяемые в данном случае лишь  $P_M$  и  $P_D$ ), может вводить в заблуждение. Например, на практике крайне редко встречаются  $P_M$  и  $P_D$  более 10-20%. Тогда оценки  $Sp$  и  $NPV$  могут принимать большие значения. При этом  $Sp$  стремится к  $(1-P_M)$ , то есть доле лиц, свободных от маркера, а  $NPV$  – к  $(1-P_D)$ , то есть доле лиц, свободных от заболевания. Из этого следует важный вывод: условные оценки информативности ( $Se$ ,  $Sp$ ,  $PPV$  и  $NPV$ ) биомаркера должны в обязательном порядке дополняться указанием общих мер сопряженности (отношения шансов или отношения рисков), а также – интегральных показателей информативности маркера (скрининговой или прогностической балансовой точности –  $SBA$  или  $PBA$ ). Только при статистической значимости сопряженности маркера и исхода,  $OR > 5,5$  и  $SBA/PBA > 0,7$  маркер является эффективным скрининговым или прогностическим классификатором.

Перечень эмпирически оцениваемых показателей информативности и сопряженности непосредственным образом определяется дизайном исследования: исследование «случай-контроль» позволяет непосредственно оценить  $Se$ ,  $Sp$ ,  $SBA$  и  $OR$ , когортное исследование –  $PPV$ ,  $NPV$ ,  $PBA$ ,  $OR$  и  $RR$ . Популяционное исследование позволяет оценить любые показатели информативности и сопряженности. Также необходимо помнить, что  $Se$ ,  $Sp$  и  $SBA$  ( $AUC$ ) характеризуют исключительно скрининговую информативность маркера.  $PPV$ ,  $NPV$  и  $PBA$  характеризует исключительно прогностическую информативность маркера.  $OR$  и  $RR$  является мерами силы сопряженности маркера и исхода.

#### *Abstract*

Physicians do not often correctly interpret the informativeness measures of biomarkers provided in scientific papers. The most common estimates are sensitivity ( $Se$ ) the probability that the biomarker is positive in a case of disease, specificity ( $Sp$ ) the probability that the marker is negative for patients who do not have a disease, the positive predictive value of a test result ( $PPV$ ) the probability of a disease in a marker-positive patient, a negative predictive value ( $NPV$ ) the probability that a marker-negative patient does not have a disease.

Does the very high  $Se$ ,  $Sp$ ,  $PPV$  or  $NPV$ , as well as the revealed a statistically significant association between marker and outcome, mean that the marker is effective? Not always so. The statistical significance of the marker-outcome association is only a necessary condition, but not a sufficient one. The practical use of the marker depends on its frequency of occurrence ( $P_M$ ) and the prevalence of the disease ( $P_D$ ) in the population under study. The fact that some conventional indices of informativeness can be high even in the absence of a real association between marker and outcome (defined in such a case only by  $P_M$  and  $P_D$ ) can be misleading. This leads us to an important conclusion: conventional indices of biomarker informativeness ( $Se$ ,  $Sp$ ,  $PPV$  and  $NPV$ ) must be supplemented by statistics that measure the strength of the relationship between marker and outcome (odds ratio or risk ratio), as well as integral measures of marker informativeness (screening or predictive balance accuracy –  $SBA$  or  $PBA$ ).

The quantity of empirically evaluated diagnostic performance measures is directly determined by the study design: the case-control study enables to measure directly  $Se$ ,  $Sp$ ,  $SBA$ , and  $OR$ , the cohort study allows to evaluate  $PPV$ ,  $NPV$ ,  $PBA$ ,  $OR$ , and  $RR$ . The population-based study provides investigators with any performance indices of diagnostic tests. It also should be kept in mind that  $Se$ ,  $Sp$  and  $SBA$  ( $AUC$ ) characterize only the screening informativeness of the marker.  $PPV$ ,  $NPV$  and  $PBA$  characterize only the predictive informativeness of the marker.  $OR$  and  $RR$  quantify the strength of the association between marker and outcome.

*Key words:* biomarker, sensitivity, specificity, positive predictive value, negative predictive value, disease prevalence, statistics

«Из объяснений я ничего не понял,  
но постепенно привык и перестал удивляться»

Понедельник начинается в субботу.  
Аркадий и Борис Стругацкие.

## Введение

На протяжении всей истории развития медицина опирается на различные диагностические признаки: будь то бинарные признаки (отсутствие или наличие симптомов, например – отеков) или количественные показатели (например, концентрация белка в моче). В эпоху современной медицины исследователи перешли от практики простого эмпирического описания частоты встречаемости симптомов при различных заболеваниях к представлению определенных оценок. В прошлые десятилетия появилось огромное количество научных работ, которые свидетельствуют о потенциальной возможности практического применения различных показателей на практике. Вместе с тем арсенал практикующего доктора значительно менее разнообразен. Для неискушенного читателя основной парадокс кроется в том, что обнаруженная статистически значимая связь маркера с определенным заболеванием и большие значения некоторых показателей информативности не гарантируют возможность эффективного применения этого маркера на практике: как правило, это необходимые, но не достаточные условия. Более того, большая часть диагностических признаков эффективна лишь для определенных целей и совершенно бесполезна для других. Простой пример: повышенная концентрация креатинина в крови является широко используемым маркером при скрининге для выявления пациентов с нарушением функции почек. Однако креатинин может длительное время оставаться в пределах нормальных значений при различных морфологических вариантах хронического гломерулонефрита, в том числе протекающих с формированием тяжелого нефротического синдрома, когда наличие клинкоморфологических признаков хронизации процесса очевидно [1]. То есть как прогностический признак концентрация креатинина в пределах референсных значений не позволяет полностью исключить заболевание почек.

Если не принимать в расчет разные способы вычисления доверительных интервалов (что останется за рамками данной работы), то можно сказать, что, в отличие от многих статистических методов, при оценке информативности биомаркеров математический аппарат чрезвычайно прост и в большинстве случаев сводится к вычислению долей. Наибольшие трудности, как правило, вызывает именно интерпретация полученных оценок. Как показывает опыт иностранных коллег, практикующие доктора нередко ошибочно интерпретируют

результаты статистического анализа (в частности – показатели информативности биомаркеров), приводимые в научных публикациях [2, 3, 4, 5, 6]. Неспособность критического осмысления приводимых оценок делает невозможным потребление нативной научной продукции. А переработанная и адаптированная научная продукция, изложенная в виде клинических рекомендаций, нередко существенно отстаёт от стремительно развивающихся представлений об оптимальных методах профилактики, диагностики и лечения различных заболеваний.

Мы приведем некоторые особенности применения наиболее широко распространенных показателей информативности тестов (чувствительность, специфичность, площадь под ROC кривой и т.д.).

## 1. Общие положения

Введем основные понятия. Под маркером мы будем подразумевать любой бинарный признак (например, *есть/нет* полулуния при морфологическом исследовании почечных биоптатов). Даже если признак количественный, в данной статье он будет условно сводиться к качественному признаку – дихотомическому суждению «есть или нет патологическое изменение», например, суточная протеинурия *больше/меньше* порогового значения. Под «заболеванием» мы будем понимать широкий спектр неблагоприятных состояний (осложнение, заболевание и т.д.).

Как правило, информативность маркеров оценивается по отношению к определенному «золотому стандарту», в качестве которого в нефрологии чаще всего (но далеко не всегда) выступает биопсия почки.

Диагностический тест чаще всего не бывает идеальным, безошибочно классифицируя субъектов с заболеванием и без заболевания: истинно положительные (*true positive, TP*) – маркер есть ( $M$ ), заболевание есть ( $D$ ) и истинно отрицательные результаты (*true negative, TN*) – маркера нет ( $\bar{M}$ ), заболевания нет ( $\bar{D}$ ). Как правило, тест дает еще неправильные классификации: ложно положительные (*false positive, FP*) – маркер есть ( $M$ ), заболевания нет ( $D$ ) и ложно отрицательные (*false negative, FN*) результаты – маркера нет ( $\bar{M}$ ), заболевание есть ( $D$ ) – рисунок 1А. При этом неопределенные результаты теста не рассматриваются и исключаются из анализа. В итоге результаты применения такого теста сводятся к таблице сопряженности  $2 \times 2$  – таблица 1.

Таблица 1 | Table 1

Таблица сопряженности для бинарного маркера и бинарного исхода  
Contingency table for the binary marker and binary outcome

	$D$ (болезнь есть)	$\bar{D}$ (болезни нет)	$\Sigma$
$M$ (маркер есть)	$TP (M, D)$	$FP (M, \bar{D})$	$TP + FP$
$\bar{M}$ (маркера нет)	$FN (\bar{M}, D)$	$TN (\bar{M}, \bar{D})$	$FN + TN$
$\Sigma$	$TP + FN$	$FP + FN$	



Рис. 1. Схематическое изображение различных вариантов сопряженности маркера и исхода (болен/здоров). Красным отмечены субъекты с маркером, черным – без маркера.

– true positive, – true negative, – false negative, – false positive.

А –  $OR \gg 1$ . Б –  $OR \approx 1$ . В –  $OR \gg 1, P_D \gg P_M$ . Г –  $OR \gg 1, P_D \ll P_M$ .  
 $P_D$  – распространенность заболевания,  $P_M$  – частота встречаемости маркера (PM)

Fig. 1. Schematic representation of various variants of conjugation of the marker and the outcome (sick/healthy). Subjects with a marker are marked in red, and subjects without a marker are marked in black.

– true positive, – true negative, – false negative, – false positive.

А –  $OR \gg 1$ . Б –  $OR \approx 1$ . В –  $OR \gg 1, P_D \gg P_M$ . Г –  $OR \gg 1, P_D \ll P_M$ .  
 $P_D$  – disease prevalence,  $P_M$  – frequency of marker occurrence

Схематично таблицу 1 можно представить на рисунке 2, где также приведены наиболее часто применяемые оценки ( $Se$  и  $Sp$ ,  $PPV$  и  $NPV$ ), а также обратные им оценки ( $FNR$  и  $FPR$ ,  $FDR$  и  $FOR$ , соответственно).

Нередко значения ячеек в этой таблице сопряженности нормируют на единицу, что при определенных условиях позволяет трактовать их как вероятности. Если выборка была получена случайным способом, а не в результате набора групп по принципу *есть заболевание/нет заболевания* (исследование случай-контроль) или по принципу *есть маркер/нет маркера* (когортное исследование), то маргинальные суммы представляют собой частоту встречаемости маркера  $P_M$  ( $TP+FP$ ) и распространенность заболевания  $P_D$  ( $TP+FN$ ). То есть можно сказать, что у случайно выбранного субъекта вероятность на-

личия маркера или заболевания соответствует  $P_M$  или  $P_D$  соответственно. Тогда, например, вероятность наличия и маркера, и заболевания  $P(M, D)$  соответствует  $TP$ , а вероятность наличия маркера при отсутствии заболевания  $P(M, \bar{D})$  –  $FP$ . Следует отметить, что на основе полей таблицы сопряженности (таблица 1) легко запомнить интерпретацию ошибок I и II рода, нередко упоминающихся в публикациях:  $FP$  соответствует вероятности ошибки I рода (найти различия там, где их нет – в данном случае: найти маркер у здоровых), а  $FN$  – вероятности ошибки второго рода (не найти различия там, где они есть – в данном случае: не выявить маркер у больных).

Несмотря на то, что в этой таблице сопряженности всего четыре поля, на ее основе можно получить большое количество оценок, каждая из которых имеет свою интерпретацию (см. таблицу 2).

## 2. Основные показатели информативности

Основные показатели информативности, варианты их названия, формулы для их вычисления, а также интерпретация оценок, описываемых в данной статье, представлены в таблице 2, а схематичное их изображение – на рисунке 2. В большинстве случаев в таблице 2 приведены формулы с использованием полей таблиц

сопряженности, отражающие смысл различных показателей. Однако мы привели также и некоторые формулы для вычисления показателей на основе оценок распространенности заболевания ( $P_D$ ) и частоты встречаемости маркера ( $P_M$ ). Основные показатели условно классифицированы в таблице 3.

**2.1. Общие показатели сопряженности маркера и заболевания** (отношение шансов и отношение рисков –  $OR$  и  $RR$ ) подробно рассмотрены нами ранее на страницах этого журнала [7]. Напомним лишь, что шансы представляют собой отношение доли субъектов с заболеванием к альтернативному варианту, то есть доле субъектов без этого заболевания среди пациентов с маркером и без маркера ( $TP/FP$  и  $FN/TN$ , соответственно). Шансы могут принимать значение от  $-\infty$  до  $+\infty$ . В свою очередь, риск представляет собой отношение доли субъектов

Таблица 2 | Table 2

**Основные показатели информативности биомаркеров**  
Main markers performance estimates

Название	Формула	Интерпретация
Распространенность заболевания, преваленс, <i>prevalence</i> ( $P_D$ )	$\frac{TP + FN}{TP + FP + TN + FN}$	Распространенность заболевания
Распространенность маркера ( $P_M$ )	$\frac{TP + FP}{TP + FP + TN + FN}$	Частота встречаемости маркера
Отношение шансов ( <i>odds ratio</i> – $OR$ )	$\frac{TP \times TN}{FP \times FN} = \frac{Se \times Sp}{(1 - Se) \times (1 - Sp)} = \frac{PPV \times NPV}{(1 - PPV) \times (1 - NPV)} = \frac{LR_+}{LR_-}$	Кратность увеличения шансов заболевания при наличии маркера
Отношение рисков ( <i>risk ratio</i> – $RR$ )	$\frac{TP \times (FN + TN)}{FN \times (TP + FP)} = \frac{PPV}{1 - NPV} = \frac{(1 - P_M) \times Se}{P_M \times (1 - Se)}$	Кратность увеличения риска заболевания при наличии маркера
Чувствительность ( <i>sensitivity</i> – $Se$ , <i>true positive rate</i> , <i>recall</i> )	$\frac{TP}{TP + FN} = \frac{PPV \times P_M}{PPV \times P_M + (1 - NPV) \times (1 - P_M)}$	Вероятность выявить маркер у больных
Специфичность ( <i>specificity</i> – $Sp$ , <i>true negative rate</i> – $TPR$ )	$\frac{TN}{TN + FP} = \frac{NPV \times (1 - P_M)}{NPV \times (1 - P_M) + (1 - PPV) \times P_M}$	Вероятность отсутствия маркера у здоровых
Частота ложноотрицательных результатов ( <i>false negative rate</i> – $FNR$ )	$\frac{FN}{FN + TP} = 1 - Se$	Вероятность отсутствия маркера у больных
Частота ложноположительных результатов ( <i>false positive rate</i> – $FPR$ )	$\frac{FP}{FP + TN} = 1 - Sp$	Вероятность выявить маркер у здоровых
Прогностическая ценность положительного результата ( <i>positive predictive value</i> – $PPV$ , <i>precision</i> )	$\frac{TP}{TP + FP} = \frac{Se \times P_D}{Se \times P_D + (1 - Sp) \times (1 - P_D)} = \frac{P_D \times LR_+}{1 - P_D + P_D \times LR_+}$	Вероятность болезни при наличии маркера
Прогностическая ценность отрицательного результата ( <i>negative predictive value</i> – $NPV$ )	$\frac{TN}{TN + FN} = \frac{Sp \times (1 - P_D)}{Sp \times (1 - P_D) + (1 - Se) \times P_D} = 1 - \frac{P_D \times LR_+}{1 - P_D + P_D \times LR_+}$	Вероятность отсутствия болезни при отсутствии маркера
Частота ложных обнаружений ( <i>false discovery rate</i> – $FDR$ )	$\frac{FP}{FP + TP} = 1 - PPV$	Вероятность отсутствия болезни при наличии маркера
Частота упущений ( <i>false omission rate</i> – $FOR$ )	$\frac{FN}{FN + TN} = 1 - NPV$	Вероятность болезни при отсутствии маркера
Скрининговая балансовая точность, площадь под ROC-кривой ( <i>screening balance accuracy</i> – $SBA$ , <i>area under ROC curve</i> – $AUC_{ROC}$ )	$\frac{Se + Sp}{2}$	Усредненная доля верно идентифицированных больных и здоровых
Прогностическая балансовая точность ( <i>predictive balance accuracy</i> – $PBA$ )	$\frac{PPV + NPV}{2}$	Усредненная доля верно идентифицированных носителей маркера и лиц без маркера
Индекс Юдена $Se - Sp$ ( <i>Youden's index</i> , $J_{Se-Sp}$ <i>statistic</i> )	$Se + Sp - 1$	Разность между долей носителей маркера среди больных и доли лиц с маркером среди здоровых
Индекс $PPV - NPV$ ( $J_{PPV-NPV}$ <i>statistic</i> )	$PPV + NPV - 1$	Разность между долей больных среди носителей маркера и доли больных среди лиц без маркера

Таблица 3 | Table 3

**Классификация показателей информативности и сопряженности маркеров**  
Classification of marker performance estimates and conjugacy of marker and outcome

Показатели	Назначение показателя	
	Скрининг (выявление лиц «группы риска» или профотбор). «Скрининговый» маркер	Подтверждение заболевания или его отсутствия. «Диагностический» маркер
Общие меры сопряженности	$RR, OR$	
Условные показатели информативности маркера	$Se, Sp, FPR, FNR$	$PPV, NPV, FDR, FOR$
Интегральные показатели информативности маркера	$SBA, J_{Se-Sp}$	$PBA, J_{PPV-NPV}$





	$D$	$\bar{D}$	Оценки	
$M$	 $TP$	 $FP$	$PPV = \frac{\text{TP}}{\text{TP} + \text{FP}}$	$FDR = \frac{\text{FP}}{\text{TP} + \text{FP}}$
$\bar{M}$	 $FN$	 $TN$	$NPV = \frac{\text{TN}}{\text{TN} + \text{FN}}$	$FOR = \frac{\text{FN}}{\text{TN} + \text{FN}}$
Оценки	$Se = \frac{\text{TP}}{\text{TP} + \text{FN}}$	$Sp = \frac{\text{TN}}{\text{TN} + \text{FP}}$	$SBA = (Se + Sp) / 2$ $J_{Se-Sp} = Se + Sp - 1$ $PBA = (PPV + NPV) / 2$ $J_{PPV-NPV} = PPV + NPV - 1$	
	$FNR = \frac{\text{FN}}{\text{TP} + \text{FN}}$	$FPR = \frac{\text{FP}}{\text{TP} + \text{FP}}$	$P_M = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$ $P_D = \frac{\text{TP} + \text{FN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$	

Рис. 2. Схематическое изображение таблицы сопряженности 2x2 (болен/здоров, маркер есть/маркера нет) и основные показатели информативности маркеров. Красным отмечены субъекты с маркером, черным – без маркера.

 – true positive,  – true negative,  – false negative,  – false positive

Fig. 2. Schematic representation of the 2x2 conjugacy table

(sick/healthy, marker present/marker not present) and the main markers performance estimates. Subjects with a marker are marked in red, and subjects without a marker are marked in black.

 – true positive,  – true negative,  – false negative,  – false positive

с заболеванием ко всем вариантам событий, то есть сумме долей субъектов с заболеванием и без заболевания среди пациентов с маркером и без маркера:  $TP / (TP + FP)$  и  $FN / (FN + TN)$ , соответственно. Поскольку риск представляет собой условную вероятность  $P(D | M)$  – вероятность болезни при условии наличия маркера или  $P(D | \bar{M})$  – вероятность болезни при условии отсутствия маркера, значение этих показателя может принимать значение от нуля до единицы. Отношение этих двух оценок в группах субъектов с маркером и без маркера дает нам  $OR$  или  $RR$ . Наиболее универсальной оценкой является  $OR$ . При этом  $OR$  является косвенной оценкой  $RR$ . Важно помнить, что  $OR$  и  $RR$  отражают силу связи маркера и заболевания.

Значения  $OR$  и  $RR$  могут лежать в интервале от нуля до  $+\infty$ . Если  $OR$  и  $RR \approx 1$ , это свидетельствует об отсутствии связи маркера и заболевания (формально – значимого изменения риска заболевания при наличии или отсутствии маркера), зна-

чения  $OR$  и  $RR > 1$  – об увеличении риска, а значения  $OR$  и  $RR < 1$  – об уменьшении риска заболевания у субъектов с маркером по отношению к субъектам без маркера. При этом  $OR$  в большинстве случаев «переоценивает»  $RR$ :  $OR > RR$ , в случае, если  $RR > 1$ , и  $OR < RR$ , если  $RR < 1$ . Значения  $OR > RR \rightarrow +\infty$  (стремятся к бесконечности) или  $OR > RR \rightarrow 0$  (стремятся к нулю) свидетельствуют о выраженном увеличении или снижении риска соответственно.

**2.2. Условные показатели информативности маркера и заболевания** наиболее распространены. Причем чаще всего в статьях приводятся значения чувствительности ( $Se$ ) и специфичности ( $Sp$ ), несколько реже – значения прогностической ценности положительного ( $PPV$ ) результата и отрицательного ( $NPV$ ) результата теста. Две пары этих показателей нередко путают.

Если внимательно посмотреть на способ вычисления этих показателей (таблица 2, рисунок 2), то нетрудно заметить, что  $Se$  представляет собой долю верно идентифицированных больных. В свою очередь  $Sp$  – это доля верно идентифицированных здоровых. Иными словами, при оценке  $Se$  объек-

том внимания являются все больные и вычисляется доля субъектов, которая была классифицирована правильно (то есть доля субъектов с маркером среди всех больных:  $TP$  из  $TP + FN$ ). При вычислении  $Sp$  объектом внимания являются все здоровые и вычисляется доля субъектов, которая была классифицирована правильно (то есть доля субъектов без маркера среди всех здоровых:  $TN$  из  $TN + FP$ ) – рисунок 2.

С практической точки зрения бывает более важно определить вероятность наличия болезни у лиц с маркером или – отсутствия болезни у лиц без маркера. Для этого существуют показатели прогностической ценности положительного результата ( $PPV$ ) и прогностической ценности отрицательного результата ( $NPV$ ) теста соответственно.  $PPV$  представляет собой долю верно идентифицированных носителей маркера, а  $NPV$  – долю верно идентифицированных лиц, свободных от маркера. Иными словами, при оценке  $PPV$  объектом внимания являются все носители маркера и вычисляется доля субъектов,

которая была классифицирована правильно (то есть доля больных среди всех субъектов с маркером:  $TP$  из  $TP+FP$ ). При оценке  $NPV$  объектом внимания являются все субъекты, свободные от маркера, и вычисляется доля субъектов, которая была классифицирована правильно (то есть доля здоровых среди всех субъектов без маркера:  $TN$  из  $TN+FN$ ) – рисунок 2.

Чтобы не перепутать интерпретации оценок  $Se$  и  $PPV$ ,  $Sp$  и  $NPV$ , необходимо помнить, что они представляют собой условные вероятности, а именно:

$Se$  – вероятность наличия маркера при условии наличия заболевания:  $P(M|D)$ ;

$Sp$  – вероятность отсутствия маркера при условии отсутствия заболевания:  $P(\bar{M}|\bar{D})$ ;

$PPV$  – вероятность наличия болезни при условии наличия маркера:  $P(D|M)$ ;

$NPV$  – вероятность отсутствия заболевания при условии отсутствия маркера:  $P(\bar{D}|\bar{M})$ .

Поскольку указанные показатели представляют собой вероятности, их значение может лежать в диапазоне от нуля до единицы.

Читателю важно помнить, что в более привычной повседневной интерпретации  $Se$  – есть та доля субъектов с заболеванием, а  $Sp$  – доля субъектов без заболевания, которые могут быть выявлены на основе значения маркера. Таким образом, эти два показателя характеризуют исключительно скрининговую эффективность маркера, то есть способность выявлять больных и здоровых, но ничего не скажут практикующему доктору о вероятности наличия или отсутствия заболевания у конкретного пациента. В свою очередь,  $PPV$  и  $NPV$  характеризуют исключительно диагностическую (= прогностическую) эффективность маркера, то есть в своей интуитивно понятной вероятностной интерпретации дадут практикующему доктору информацию о вероятности наличия или отсутствия заболевания у конкретного пациента при наличии или отсутствии маркера соответственно, но ничего не скажут о доле больных и здоровых лиц, которые могут быть выявлены на основе значения маркера.

Существуют также и обратные оценки по отношению к  $Se$  и  $Sp$ ,  $PPV$  и  $NPV$  –  $FNR$  и  $FPR$ ,  $FDR$  и  $FOR$  соответственно – см. таблицу 2 и рисунок 2.  $FNR$  ( $1-Se$ ) представляет собой долю неверно идентифицированных больных (поскольку на основании значения маркера они отнесены к здоровым), то есть долю субъектов без маркера среди всех больных:  $FN$  из  $TP+FN$ . Это важный показатель – упущенные при скрининге больные.  $FPR$  ( $1-Sp$ ) представляет собой долю неверно идентифицированных здоровых (поскольку на основании значения маркера они отнесены к больным), то есть долю субъектов с маркером среди всех здоровых:  $FP$  из  $FP+TN$ .

В свою очередь,  $FDR$  ( $1-PPV$ ) представляет собой долю неверно идентифицированных носи-

телей маркера, то есть долю здоровых среди всех носителей маркера (поскольку на основании значения маркера они могут считаться больными):  $FP$  из  $FP+TP$ , а  $FOR$  ( $1-NPV$ ) – долю неверно идентифицированных субъектов без маркера (поскольку на основании значения маркера они могут считаться больными), то есть долю больных среди лиц без маркера:  $FN$  из  $FN+TN$ .

### 3. Практические аспекты интерпретации показателей информативности маркеров

Потребность в практическом применении полученной из публикаций информации нередко порождает ряд важных вопросов:

1. Означает ли, что большое значение  $Se$ ,  $Sp$ ,  $PPV$  или  $NPV$  будет свидетельствовать, что маркер эффективен? Далеко не всегда! Эти показатели могут достигать больших значений даже при полном отсутствии реальной сопряженности с заболеванием. Справедливо и обратное.
2. Означает ли, что маркер, способный с высокой эффективностью выявлять больных при скрининге, будет эффективен и для подтверждения диагноза? Чаще всего бывает наоборот. Помимо силы сопряженности маркера и заболевания (в качестве меры которой мы будем использовать главным образом  $OR$ ), на эффективность применения маркера оказывают непосредственное влияние  $P_D$  и  $P_M$ .
3. Как оценить среднюю скрининговую или диагностическую ценность теста? Для этого целесообразно использовать интегральные показатели информативности маркера.
4. Если практикующего доктора в первую очередь интересует вероятность наличия заболевания при положительном значении маркера, а не доля лиц, которая может быть выявлена в результате использования теста, почему авторы публикаций проводят оценки  $Se$  и  $Sp$ , но не  $PPV$  и  $NPV$ ? Потому что перечень эмпирически оцениваемых показателей информативности и сопряженности непосредственным образом определяется дизайном исследования.

Рассмотрим эти вопросы и дадим на них подробные ответы.

#### 3.1. Ловушка изолированной оценки условных показателей информативности маркера.

Сначала для лучшего понимания рассмотрим предельные варианты сопряженности маркера и заболевания. В редких и порой сомнительных случаях, когда маркер чрезвычайно сильно ассоциирован с заболеванием ( $OR$ ,  $RR \rightarrow +\infty$ ), тест чрезвычайно эффективен, «идеальный маркер»: он позволяет с высокой долей вероятности отличить больных и здоровых. В этом случае:

- все или почти все больные имеют маркер, а здоровые – свободны от маркера:  $Se$ ,  $Sp \rightarrow 1$ ;

- все или почти все носители маркера больны, а лица, свободные от маркера – здоровы:  $PPV, NPV \rightarrow 1$ ;
- частота неверных классификаций стремится к нулю:  $FNR, FPR, FDR, FOR \rightarrow 0$ .

**Пример 1.** Примером такого почти идеального маркера может быть концентрация Lyso-Gb<sub>3</sub> (глоботриаозилсфингозина) в моче у пациентов с болезнью Фабри. Однако чувствительность этого маркера снижается у женщин, а также при неклассической форме заболевания, в связи с чем концентрация Lyso-Gb<sub>3</sub> оценивалась совместно с аналогами – гликофосфолипидами, структурно сходными с Lyso-Gb<sub>3</sub>, но имеющими различия в сфингозиновой части молекул. Значения суммарной концентрации  $>0,25$  пмоль на ммоль креатинина, обеспечивают  $Se = Sp = AUC = 1, OR \rightarrow \infty$ . При этом  $P_M = P_D$ , этот маркер встречается только у пациентов с болезнью Фабри и не встречается в иных случаях [8].

Обратная ситуация гораздо менее однозначна. При  $OR, RR \rightarrow 1$  тест будет неэффективен. Несмотря на отсутствие значимой сопряженности маркера и заболевания, которая определяет возможность практического применения маркера, тест все равно будет давать некоторую долю верных классификаций, определяемых  $P_M$  и  $P_D$ . То есть, как видно из рисунка 1Б, многие больные являются носителями маркера ( $TP$ ), а многие здоровые субъекты свободны от маркера ( $TN$ ). Однако в отличие от первого варианта ( $OR, RR \rightarrow +\infty$ ) тест будет давать и схожую долю неверных классификаций ( $FP$  – значительная доля здоровых также являются носителями маркера,  $FN$  – значительная доля больных свободны от маркера). А именно:

- маркер можно будет обнаружить у больных и здоровых с примерно одинаковой частотой, которая будет стремиться к его популяционной частоте ( $Se, FPR \rightarrow P_M$ );
- вероятность болезни будет примерно одинакова у носителей маркера и лиц, свободных от него, и будет стремиться к распространенности заболевания:  $PPV, FOR \rightarrow P_D$ ;
- маркер будет отсутствовать у больных и здоровых с примерно одинаковой частотой, которая также будет определяться его популяционной частотой:  $Sp, FNR \rightarrow 1 - P_M$ ;
- вероятность отсутствия болезни будет примерно одинакова у носителей маркера и лиц, свободных от него, и также будет определяться распространенностью заболевания:  $NPV, FDR \rightarrow 1 - P_D$ .

Нетрудно заметить, что вероятность выявить маркер определяется главным образом  $P_M$ , а вероятность выявить больного –  $P_D$ . Иными словами, чем чаще встречается маркер, тем больше будет значение  $Se$  (верные классификации), но одновременно более-менее пропорционально увеличивается и  $FPR$ . То есть, вероятность выявить маркер увеличивается по мере увеличения частоты его встречаемости ( $P_M$ )

вне зависимости от того, болен субъект или нет. В свою очередь, чем чаще встречается заболевание, тем больше значение  $PPV$ , но одновременно увеличивается и значение  $FOR$ . То есть, вероятность заболевания определяется главным образом его распространенностью ( $P_D$ ) вне зависимости от значения маркера.

Тот факт, что некоторые условные показатели информативности могут принимать большие значения даже в условиях отсутствия реальной сопряженности маркера (определяемые в данном случае лишь распространенностью маркера и заболевания, но не  $OR$  и  $RR$ , при  $OR, RR \rightarrow 1$ ), может вводить в заблуждение. На практике мы крайне редко имеем дело с  $P_M$  и  $P_D$  более 10-20%. Тогда оценки  $Sp$  и  $NPV$  могут принимать большие значения. При этом  $Sp \rightarrow 1 - P_M$ , а  $NPV \rightarrow 1 - P_D$ . Например, означает ли, что маркер с  $Se \rightarrow P_M \approx 0,1$  и  $Sp \rightarrow 1 - P_M \approx 0,9$  может быть с успехом применен при профотборе для выявления здоровых лиц? Нет, потому что тест будет давать и большую долю ложных срабатываний ( $FNR \approx 0,9$  при  $FPR \approx 0,1$ ).

Из этого следует важный вывод: условные оценки информативности биомаркера должны в обязательном порядке дополняться указанием общих мер сопряженности ( $OR, RR$ ), а также – интегральных показателей информативности маркера ( $SBA$  или  $PBA$ ).

**3.2. Назначение маркера: скрининг или диагностика?** На особенности практического применения маркера непосредственное влияние оказывают частота его встречаемости ( $P_M$ ) и распространенность заболевания ( $P_D$ ) в анализируемой популяции. В некоторых случаях маркер встречается примерно с той же частотой, что и заболевание (то есть  $P_M \approx P_D$ ), тогда возможность его использования определяется главным образом силой сопряженности ( $OR, RR$ ) – рисунок 1А. А именно: при  $OR, RR >> 1$  (значительно больше единицы) большие значения  $Se$  или  $Sp$  определяют возможность использования маркера для скрининга (для выявления больных или здоровых, соответственно), а большие значения  $PPV$  и  $NPV$  – для диагностики (для подтверждения заболевания или его отсутствия).

Если  $P_M \approx P_D$ , а  $OR, RR >> 1$  (как в примере с болезнью Фабри – *пример 1*), это означает, что маркер можно выявить только или почти только у субъектов с изучаемым заболеванием, и он не встречается у субъектов без заболевания. Такой маркер будет высоко эффективен и для скрининга, и для подтверждения наличия или отсутствия заболевания при наличии или отсутствии маркера ( $Se, Sp, PPV, NPV \rightarrow 1$ ). Всегда ли при  $P_M \approx P_D$  маркер так хорош? Конечно, нет! Даже если  $P_M \approx P_D$ , маркер может встречаться и у некоторых субъектов с заболеванием и отсутствовать у некоторых субъектов без заболевания. На рисунке 1Б  $P_M \approx P_D$ , однако маркер бесполезен на практике и для скрининга, и для диагностики, поскольку значимая сопряженность



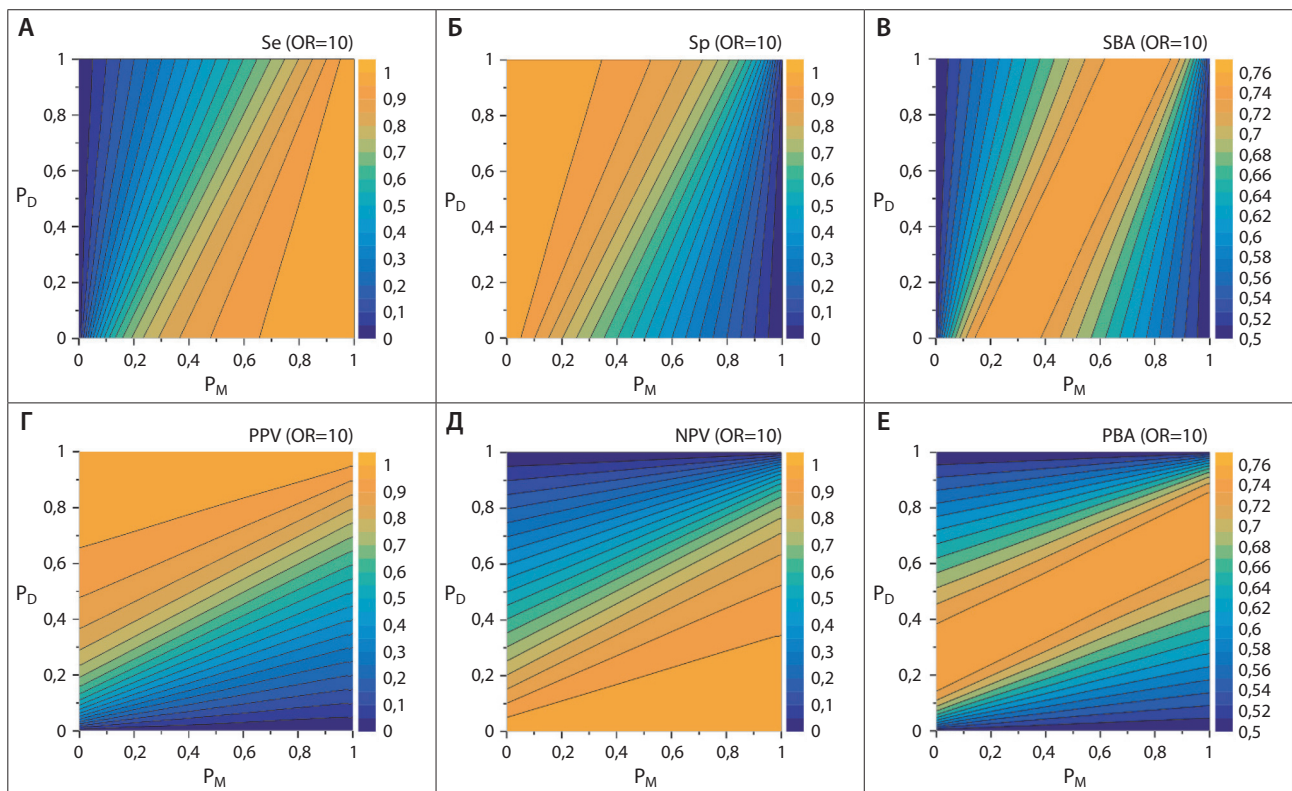


Рис. 3. Показатели информативности маркера при различных значениях распространенности заболевания ( $P_D$ ) и частоты встречаемости маркера ( $P_M$ ) при  $OR=10$ . Искусственный датасет

Fig. 3. Markers performance estimates and outcome at different values of the disease prevalence ( $P_D$ ) and the frequency of marker occurrence ( $P_M$ ) when  $OR=10$ . Artificial dataset

его с заболеванием отсутствует:  $OR, RR \rightarrow 1$ . И чем меньше сопряженность маркера и заболевания, тем реже он встречается у субъектов с заболеванием и все чаще – у субъектов без заболевания. То есть, по мере приближения  $OR, RR$  к единице уменьшаются доли верных классификаций ( $Se, Sp, PPV, NPV$ ) и возрастает доля неверных ( $FNR, FPR, FDR, FOR$ ).

Заболевание и маркер далеко не всегда распределены в популяции с примерно равной частотой. Обратимся к теории. Для наглядного представления влияния распространенности заболевания и частоты встречаемости маркера на показатели информативности и сопряженности мы смоделировали датасет (*data set*). Результаты симуляции представлены на рисунке 3. Сила сопряженности маркера и заболевания при любых  $P_D$  и  $P_M$  была фиксирована на значении  $OR=10$  (отметим, что при больших значениях  $OR$  «желтые» области будут прибывать, а «синие» – убывать и наоборот).

Чувствительность и специфичность представляют собой вероятность наличия маркера у больного и его отсутствия у здорового, соответственно. Поэтому при фиксированном значении  $OR$   $Se$  и  $Sp$  главным образом зависят от  $P_M$  и в меньшей степени – от  $P_D$  (рисунок 3 А и Б). Закономерно, что чувствительность возрастает при увеличении частоты встречаемости маркера (особенно для редких

заболеваний) и постепенно убывает при увеличении распространенности заболевания. Специфичность имеет обратную зависимость от  $P_M$  и  $P_D$ . Иными словами, чем чаще встречается маркер, тем эффективнее он будет при скрининге для выявления лиц с заболеванием (то есть, при фиксированном значении  $OR, RR, P_D$  и  $P_M \rightarrow 1, Se \rightarrow 1, FNR \rightarrow 0$ ). Вместе с тем, в этих же условиях по мере увеличения частоты встречаемости маркера будет увеличиваться вероятность наличия маркера у здоровых ( $FPR \rightarrow 1$ ) и уменьшаться специфичность ( $Sp \rightarrow 0$ ).

$PPV$  и  $NPV$  представляют собой вероятность заболевания при наличии маркера и вероятность отсутствия заболевания при отсутствии маркера соответственно. Поэтому  $PPV$  и  $NPV$  главным образом зависят от  $P_D$  и в меньшей степени – от  $P_M$  (рисунок 3 Г и Д).  $PPV$  возрастает при увеличении распространенности заболевания (особенно в случае редкого маркера) и постепенно убывает при увеличении  $P_M$ .  $NPV$  имеет обратную зависимость от  $P_M$  и  $P_D$ . Иными словами, чем чаще встречается заболевание, тем с большей вероятностью маркер будет свидетельствовать о его наличии (то есть при фиксированном значении  $OR, RR, P_M$  и  $P_D \rightarrow 1, PPV \rightarrow 1, FDR \rightarrow 0$ ). Вместе с тем, в этих же условиях по мере увеличения частоты встречаемости заболевания будет увеличиваться его вероятность при отрица-

тельном значении маркера ( $FOR \rightarrow 1$ ) и уменьшаться  $NPV$  ( $NPV \rightarrow 0$ ).

Вернёмся от теории к практике. То же самое можно схематично представить на рисунках 1 В и Г. В случае широко распространенного заболевания и редко встречающегося маркера ( $P_D \gg P_M$ ) и  $OR$ ,  $RR \gg 1$  (рисунках 1 В):

- $Se \rightarrow 0$ , то есть маркер плохо подходит для скрининга с целью выявления больных, поскольку он упустит большую долю больных ( $FNR \rightarrow 1$ );
- $Sp \rightarrow 1$ , то есть маркер хорошо подходит для скрининга с целью выявления здоровых, поскольку доля носителей маркера среди здоровых очень мала ( $FPR \rightarrow 0$ );
- $PPV \rightarrow 1$ , то есть положительное значение маркера с большой вероятностью свидетельствует о наличии заболевания, при этом доля здоровых среди носителей маркера очень мала ( $FDR \rightarrow 0$ );
- $NPV \rightarrow 0$ , то есть при отсутствии маркера нельзя сделать заключение, что субъект здоров с высокой вероятностью (вероятность отсутствия заболевания при отсутствии маркера низка), при этом доля больных среди субъектов без маркера велика ( $FOR \rightarrow 1$ ).

Таким образом, при  $P_D \gg P_M$  и  $OR$ ,  $RR \gg 1$  маркер может быть эффективен при скрининге для выявления группы здоровых лиц (например – проведение профотбора) или для подтверждения заболевания.

**Пример 2.** Примером такого маркера могут быть антитела к аннексину А5 при диагностике антифосфолипидного синдрома (АФС). Пациенты с первичным и вторичным АФС значительно чаще имеют антитела к аннексину А5 класса IgM, чем пациенты с заболеваниями/состояниями, клиническая картина которых позволяет заподозрить АФС (не АФС-ассоциированные тромбозы, системная красная волчанка без АФС и не-АФС заболевания, связанные с беременностью). Авторы публикации [9] получили следующую таблицу сопряженности:

IgM, ед./мл	D (АФС)	$\bar{D}$ (АФС нет)	$\Sigma$
$M (>18)$	26	1	27
$\bar{M} (\leq 18)$	144	103	247
$\Sigma$	170	104	274

В этом случае  $P_D \gg P_M$  ( $P_D/P_M=0,62/0,099 \approx 6$ ). Авторы получили следующие оценки:  $Se=0,153$ ,  $Sp=0,99$ ,  $PPV=0,963$ ,  $NPV=0,417$ . Маркер сильно сопряжен с заболеванием (наличием АФС):  $OR=18,6$ ,  $RR=1,65$ ,  $p<0,0001$  (отметим, что  $OR \gg RR$ , что весьма характерно для большой распространенности заболевания в исследуемой популяции [7]). Таким образом, маркер позволит выявить 99% лиц без АФС (маркер отсутствует почти у всех субъектов без АФС: 103 из 104) и всего 15,3% лиц с АФС (маркер присутствует у 26 из 144 субъектов). При этом положительное значение маркера (IgM > 18 ед./мл)

с вероятностью 96,3% свидетельствует в пользу АФС (26 из 27 субъектов в этом исследовании). Если же значение IgM  $\leq 18$  ед./мл, то вероятность отсутствия АФС всего 41,7% (103 из 247 пациентов). То есть, для исключения АФС необходимо применение дополнительного диагностического признака.

В случае редко встречающегося заболевания и широко распространенного маркера ( $P_D \ll P_M$ ) и  $OR$ ,  $RR \gg 1$  (рисунках 1 Г):

- $Se \rightarrow 1$ , то есть маркер хорошо подходит для скрининга с целью выявления больных, доля больных без маркера крайне низка ( $FNR \rightarrow 0$ );
- $Sp \rightarrow 0$ , то есть маркер плохо подходит для скрининга с целью выявления здоровых, поскольку значительная доля здоровых лиц имеют маркер ( $FPR \rightarrow 1$ );
- $PPV \rightarrow 0$ , то есть при наличии маркера нельзя сделать заключение, что субъект болен с высокой вероятностью (вероятность заболевания при наличии маркера низка), при этом доля здоровых среди носителей маркера велика ( $FDR \rightarrow 1$ );
- $NPV \rightarrow 1$ , то есть отрицательное значение маркера с большой вероятностью свидетельствует об отсутствии заболевания, при этом доля больных среди субъектов без маркера мала ( $FOR \rightarrow 0$ ).

Таким образом, при  $P_D \ll P_M$  и  $OR$ ,  $RR \gg 1$  маркер может быть эффективен при скрининге для выявления группы больных (например – онкоскрининг) или исключения заболевания.

**Пример 3.** Примером такого маркера может быть нейтрофильный желатиназо-ассоциированный липокалин (NGAL) при оценке риска развития острого почечного повреждения (ОПП, под которым авторы понимали увеличение сывороточного креатинина более чем на 0,5 мг/дл или потребность в заместительной почечной терапии) у пациентов с подозрением на сепсис (в соответствии с критериями Sepsis-2 [10]) в отделениях интенсивной терапии [11]. При повышении NGAL > 150 нг/л (маркер) риск ОПП (заболевание) существенно возростал. Авторы получили следующую матрицу:

NGAL, нг/л	D (ОПП)	$\bar{D}$ (ОПП нет)	$\Sigma$
$M (>150)$	23	310	333
$\bar{M} (\leq 150)$	1	327	328
$\Sigma$	24	637	661

В этом случае  $P_D \ll P_M$  ( $P_M/P_D=0,504/0,036 \approx 14$ ). Авторы получили следующие оценки:  $Se=0,958$ ,  $Sp=0,513$ , имея таблицу сопряженности нетрудно получить и оценки  $PPV=0,069$ ,  $NPV=0,997$  (дизайн исследования позволяет это – см. ниже). При этом  $OR=24,26$ ,  $RR=22,66$ ,  $p<0,0001$  (отметим, что  $OR \approx RR$  и напомним, что это весьма характерно для  $P_D < 10\%$  [7]). Таким образом, маркер может быть эффективен в качестве скринингового признака для выявления лиц с повышенным риском развития ОПП – позволит выявить 95,8% больных,

у которых течение сепсиса осложнится развитием ОПП в течение 72 часов (маркер присутствует почти у всех пациентов с ОПП: 23 из 24 пациентов) и 51,3% пациентов, у которых ОПП не разовьется (327 из 637 пациентов). При этом положительное значение маркера ( $NGAL > 150$  нг/л) лишь с вероятностью 6,9% свидетельствует в пользу развития ОПП у пациента (ОПП развилось у 23 пациентов из 333 пациентов с маркером). То есть, для подтверждения ОПП необходимо применение дополнительного диагностического признака. Если же значение  $NGAL \leq 150$  нг/л, то вероятность того, что у пациента не будет ОПП – 99,7% (327 из 328 пациентов).

**3.3. Интегральные показатели информативности маркера.** Из рисунков 3 А и Б очевидно, что по мере возрастания  $Se$  убывает  $Sp$ , а, как следует из рисунков 3 Г и Д, по мере увеличения  $PPV$  убывает  $NPV$ . При этом с практической точки зрения важно, чтобы оба этих показателя принимали значения, близкие к единице, что бывает далеко не всегда. Однако относительное несовершенство маркера не означает невозможность его применения на практике (как мы показали на примерах выше, маркер может применяться главным образом для выявления больных или здоровых лиц при скрининге, подтверждения или исключения заболевания) при диагностике. Важно получить усреднённую оценку верных классификаций: доли носителей маркера среди больных и лиц, свободных от маркера, среди здоровых ( $Se$  и  $Sp$  соответственно), или же доли больных среди носителей маркера и доли здоровых среди лиц, свободных от маркера ( $PPV$  и  $NPV$ , соответственно). Усредненный показатель  $(Se+Sp)/2$ , представляющий собой скрининговую балансовую точность (*screening balance accuracy* –  $SB_A$ ), чаще всего именуют площадью под ROC кривой (*area under ROC curve* –  $AUC_{ROC}$ ), а усредненный показатель  $(PPV+NPV)/2$  можно обозначить как прогностическую балансовую точность (*predictive balance accuracy* –  $PB_A$ ). Поскольку в данном случае мы обсуждаем этот показатель в отрыве от ROC-анализа, вместо  $AUC_{ROC}$  мы будем использовать термин  $SB_A$ . Существует множество интерпретаций этих показателей, наиболее простыми из которых являются: усреднённая доля верно идентифицированных больных и здоровых (для  $SB_A$ ) и усреднённая доля верно идентифицированных носителей маркера и лиц без маркера. Очевидно, что интерпретации диагностической и прогностической балансовой точности очень похожи. Тем не менее не следует забывать, что эти показатели вычисляются на основе разных оценок и характеризуют разные аспекты информативности маркера. Как видно на рисунке 3 В и Е,  $SB_A$  более критично зависит от распространенности маркера, а  $PB_A$  – от распространенности заболевания.

Исходя из определения  $SB_A$  и  $PB_A$  нетрудно заметить, что значение этих показателей, близкое 0,5, свидетельствует о том, что маркер бесполезен – то

есть, решение, принятое на его основе, близко к случайному. А именно: маркер может обладать большим значением  $Sp$  (то есть, большая часть здоровых лиц не имеют маркера), при этом значение  $Se$  будет небольшим (то есть большая часть больных также не имеют маркера). Значения  $SB_A$  и  $PB_A$  можно условно ранжировать следующим образом:

0,5–0,6 – заведомо плохой классификатор,  
0,6–0,7 – низкая эффективность классификатора,  
0,7–0,8 – средняя эффективность классификатора,  
0,8–0,9 – высокая эффективность классификатора,  
>0,9 – отличный классификатор.

Значения  $SB_A$  и  $PB_A$  менее 0,5 свидетельствуют о том, что маркер работает с точностью до обратного.

Важным аспектом является то, что оценки  $Se$  и  $Sp$ ,  $PPV$  и  $NPV$  могут принимать значения от нуля до единицы в зависимости от значений  $P_M$  и  $P_D$  (это отчетливо видно на рисунке 3), в то время как максимальное значение  $SB_A$  и  $PB_A$  определяется  $OR$  (то есть, силой сопряженности маркера и заболевания) [12]:

$$SB_{A_{max}} = PB_{A_{max}} = \frac{\sqrt{OR}}{\sqrt{OR} + 1}$$

Исходя из формулы, приведенной выше, трудно получить минимальное значение  $OR$ , которое необходимо для достижения приемлемой для возможности практического применения информативности маркера  $SB_{A_{min}} = 0,7$ , а именно  $OR_{min} = 5,5$ . При меньших значениях  $OR$  маркер априори будет обладать низкой информативностью. Вместе с тем, даже при больших значениях  $OR$  маркер может обладать недостаточной информативностью. Как следует из рисунка 3, максимальные значения  $SB_A$  и  $PB_A$  достижимы только при определенных условиях, а именно – определенных значениях  $P_M$  и  $P_D$  (рисунок 3 В и Е). То есть, большое значение  $OR$  не означает большое значение  $SB_A$  и  $PB_A$ . Из этого следует важный вывод: условные оценки информативности биомаркера должны в обязательном порядке дополняться не только указанием общих ( $OR$ ,  $RR$ ) мер сопряженности, но и интегральными показателями информативности ( $SB_A$  и/или  $PB_A$  в зависимости от дизайна исследования).

В *примере 2* при  $OR = 18,6$  и  $Se = 0,153$ , а  $Sp = 0,99$ ,  $SB_A = 0,572$ , то есть, несмотря на то, что маркер очень сильно сопряжен с заболеванием, на практике он плохой классификатор и при скрининге, в среднем работает плохо: он не позволяет эффективно выявлять и больных, и здоровых ( $SB_A = 0,572$  при  $Se = 0,153$ ). Он может быть применён только для выявления здоровых лиц, например, при профотборе ( $Sp = 0,99$ ). При этом как диагностический признак маркер работает значительно лучше: при  $PPV = 0,963$ ,  $NPV = 0,417$ ,  $PB_A = 0,69$  – маркер обладает приемлемой эффективностью.

В *примере 3*, наоборот, при  $OR = 24,26$ ,  $Se = 0,958$  и  $Sp = 0,513$ ,  $SB_A = 0,736$ , то есть, маркер обладает

приемлемой эффективностью при скрининге. В то же время, он является плохим классификатором при использовании его в диагностике:  $PPV=0,069$ ,  $NPV=0,997$ ,  $PBA=0,533$ . Отсутствие маркера свидетельствует об отсутствии заболевания с вероятностью 99,7%, в то время как при его наличии вероятность заболевания всего 6,9%.

При таких наблюдаемых значениях  $OR$  (18,6 и 24,26),  $SBA_{max}=PBA_{max}=0,812$  в примере 2 и  $SBA_{max}=PBA_{max}=0,831$  – в примере 3. Однако в реальности (в примерах 2 и 3) мы наблюдаем совсем иную картину (см. выше) при этом  $SBA \neq PBA$ . Это вполне закономерно, поскольку максимальные значения  $SBA$  и  $PBA$  достижимы только при определенных условиях, а именно определенных значениях  $P_D$  и  $P_M$ .

Кроме этого обратим внимание читателя, что в примере 2 небольшое значение  $SBA$  (0,572), а в примере 3 небольшое значение  $PBA$  (0,533) авторы наблюдали при убедительной статистической значимости связи маркера и заболевания ( $p < 0,0001$  в обоих случаях). Отсюда следует важный вывод: статистически значимая связь маркера и заболевания – это необходимое, но недостаточное условие для высокой эффективности маркера как классификатора (скринингового или прогностического). При анализе публикаций необходимо четко осознавать, насколько маркер применим на практике и каковы особенности его применения.

Другой распространённой интегральной оценкой эффективности теста, вычисляемой на основе  $Se$  и  $Sp$ , является индекс Юдена (Youden's J statistic). Этот показатель является разностью между долей верно идентифицированных больных ( $TPR$ ,  $Se$ ) и долей неверно идентифицированных здоровых ( $FPR$ ), то есть, доли носителей маркера среди больных  $P(M|D)$  и доли лиц с маркером среди здоровых  $P(M|\bar{D})$ . Посмотрев на таблицу 2, нетрудно заметить, что

$$J_{Se-Sp} = TPR - FPR = TRP - (1 - TNR) =$$

$$= TPR - 1 + TNR = Se + Sp - 1.$$

Индекс Юдена нередко приводится в публикациях [13, 14, 15], главным образом, в контексте ROC-анализа (этот аспект будет рассмотрен нами в следующей публикации). В данном случае необходимо запомнить, что этот показатель можно рассматривать как интегральный показатель информативности маркера (наравне с  $SBA$ ).

Подобный индекс можно рассчитать и для  $PPV$  и  $NPV$ . Этот показатель в свою очередь является разностью между долей верно идентифицированных носителей маркера ( $PPV$ ) и неверно идентифицированных лиц, свободных от маркера ( $FOR$ ), то есть доли больных среди носителей маркера  $P(D|M)$  и доли больных среди лиц без маркера  $P(D|\bar{M})$ :

$$J_{PPV-NPV} = PPV - FOR = PPV - (1 - NPV) =$$

$$= PPV + NPV - 1.$$

Индекс Юдена формально может принимать значения от нуля до единицы, однако максимально

возможное его значение можно вычислить исходя из значения  $OR$  [12]:

$$J_{Se-Sp_{max}} = J_{PPV-NPV_{max}} = \frac{\sqrt{OR} - 1}{\sqrt{OR} + 1}$$

Зависимость индекса Юдена от  $P_M$  и  $P_D$  очень похожа на зависимость  $SBA$  и  $PBA$ .

**3.4. Показатели информативности и дизайн исследования.** Дизайн исследования, а точнее, способ формирования выборки, определяют возможность оценки различных аспектов сопряженности маркера и заболевания. Этот факт необходимо обязательно учитывать при планировании исследования, если в нем планируется оценить информативность маркера. Кроме того, нельзя забывать об этом и при критическом анализе публикаций, где описывается информативность биомаркеров.

На рисунке 2 показатели информативности сгруппированы неслучайно. Очевидно, что показатели  $Se$  и  $Sp$  вычисляются по столбцам, а  $PPV$  и  $NPV$  – по строкам. При дизайне исследования «случай-контроль» группы формируются по принципу «заболевание есть/заболевание нет». Как правило, при таком дизайне соотношения групп далеки от естественных. Таким образом, автор формирует группы с необходимым ему соотношением, исходя из целей исследования. Во-первых, это делает невозможной эмпирическую оценку распространённости заболевания ( $P_D$ ). Во-вторых, поскольку автор произвольно задает соотношения маргинальных сумм по столбцам  $D/\bar{D}$  (на рисунке 2 – соотношение количества больных к здоровым), непосредственная оценка показателей информативности, вычисляемых по строкам (то есть,  $PPV$  и  $NPV$  и, соответственно,  $PBA$  и  $J_{PPV-NPV}$ ) в рамках исследования невозможна. Такое исследование позволит непосредственно оценить только показатели информативности, вычисляемые по столбцам:  $Se$  и  $Sp$  и, соответственно,  $SBA$  и  $J_{Se-Sp}$ .

Если это выразить в виде формул, то пусть  $k$  – заданная исследователем доля лиц с заболеванием, тогда  $1-k$  – доля лиц без заболевания. При расчете, например,  $Se$  получим:

$$Se = \frac{TP \times k}{TP \times k + FN \times k}$$

Очевидно, что  $k$  в знаменателе можно смело вынести за скобки и сократить. Значение  $Se$  не зависит от  $k$ .

При расчете  $PPV$  получим:

$$PPV = \frac{TP \times k}{TP \times k + FP \times (1 - k)}$$

В данном случае значение показателя  $PPV$  будет сильно зависеть от  $k$ , произвольно выбираемого автором.

При когортном исследовании автором формируются группы пациентов с маркером и без маркера, соотношение этих групп также выбирается автором произвольно. То есть, автор произвольно устанавливает соотношение маргинальных сумм по строкам. Это делает невозможной непосредственную оценку

показателей, вычисляемых по столбцам, то есть,  $Se$  и  $Sp$  и, соответственно,  $SBА$  и  $J_{Se-Sp}$ . При когортном исследовании возможно оценить только показатели информативности, вычисляемые по строкам, то есть,  $PPV$  и  $NPV$  и, соответственно,  $PBA$  и  $J_{PPV-NPV}$ .

Напомним, что исследование «случай-контроль» не позволяет непосредственно оценить  $RR$ . При этом показатель  $OR$  является универсальным показателем сопряженности и может быть оценен при любом дизайне исследования. Данный вопрос подробно, с примерами был рассмотрен нами ранее [7].

Возможен еще один вариант формирования выборки: субъекты включаются в исследование без учета того, больны они или здоровы, являются носителями маркера или свободны от него. В случае случайного формирования выборки (популяционное исследование) и ее репрезентативности, долю больных из общего пула субъектов можно рассматривать как распространенность заболевания, также называемую преваленсом (prevalence) –  $P_D$ . В терминах вероятности этот показатель можно обозначить как  $P(D)$  – безусловную (маргинальную) вероятность наличия заболевания. В свою очередь, доля носителей маркера из общего пула субъектов отражает его распространённость,  $P(M)$  – безусловную вероятность наличия маркера. Аналогично могут быть получены показатели  $P(\bar{D})$  – безусловной вероятности отсутствия заболевания и  $P(\bar{M})$  – безусловной вероятности отсутствия маркера.

Читателю в обязательном порядке следует помнить о том, что возможность адекватной оценки показателей информативности и сопряженности определяется дизайном исследования.

Отметим, что, обладая точными знаниями о  $P_M$  и  $P_D$ , можно конвертировать различные оценки (некоторые формулы для конверсии приведены в таблице 2). Однако данные о распространенности заболевания и частоте встречаемости маркера могут быть подвержены существенной вариабельности, что снижает точность полученных оценок. В связи с этим такая конверсия выполняется нечасто. Вместе с тем, конверсию различных показателей информативности можно встретить в мета-анализах [16, 17].

### Заключение

Вместо заключения позволим себе сформулировать основные тезисы этой статьи:

1. Маркер может обладать высокими показателями информативности даже при отсутствии реальной сопряженности с изучаемым заболеванием.
2. Условные показатели сопряженности ( $Se$  и  $Sp$ ,  $PPV$  и  $NPV$ ) должны дополняться указанием интегральных показателей информативности ( $SBА$  или  $PBA$  соответственно) и показателей сопряженности (чаще –  $OR$ , реже –  $RR$ ).
3. Маркеры с  $OR < 5,5$  обладают низкой информативностью (значения  $SBА$  и  $PBA < 0,7$ ).

4. На информативность и особенности практического использования маркера непосредственное влияние оказывают частота его встречаемости и распространенность заболевания в исследуемой популяции.
5. При  $OR, RR \gg 1$  большие значения  $Se$  или  $Sp$  определяют возможность использования маркера для скрининга (для выявления больных или здоровых соответственно), а большие значения  $PPV$  и  $NPV$  – для диагностики (для подтверждения заболевания или его отсутствия).
6. В случае если  $OR, RR \gg 1$ , а заболевание встречается значительно чаще маркера ( $P_D \gg P_M$ ), маркер может быть эффективен при скрининге для выявления здоровых лиц (почти все здоровые свободны от маркера) и для подтверждения заболевания (почти все субъекты с маркером больны). Однако этот маркер не подходит для скрининга с целью выявления лиц с заболеванием, так как значительная доля больных свободны от маркера, а также маркер не подходит для исключения заболевания, т.к. значительная доля субъектов без маркера больны.
7. В случае если  $OR, RR \gg 1$ , а заболевание встречается значительно реже маркера ( $P_D \ll P_M$ ), маркер может быть эффективен при скрининге для выявления больных (почти все больные имеют маркер) и для исключения заболевания (почти все субъекты, свободные от маркера, здоровы). Однако этот маркер не подходит для скрининга с целью здоровых лиц, так как значительная доля здоровых также имеют маркер, а также не подходит для подтверждения заболевания, так как значительная доля носителей маркера здоровы.
8. Исследование «случай-контроль» позволяет непосредственно оценить  $Se$  и  $Sp$  и, соответственно,  $SBА$  и  $J_{Se-Sp}$ ,  $OR$ ; когортное исследование –  $PPV$  и  $NPV$  и, соответственно,  $PBA$  и  $J_{PPV-NPV}$ , а также  $OR$  и  $RR$ . Популяционное исследование позволяет оценить любые показатели информативности и сопряженности.

*Авторы не имеют конфликта интересов*

*The authors declare no conflict of interest*

### Список литературы

1. Kidney Disease: Improving Global Outcomes (KDIGO) Glomerulonephritis Work Group. KDIGO Clinical Practice Guideline for Glomerulonephritis. *Kidney inter., Suppl.* 2012; 2: 139-274. DOI: 10.1038/kisup.2012.9
2. *Puhan MA, Steurer J, Bachmann LM, ter Riet G.* A randomized trial of ways to describe test accuracy: the effect on physicians' post-test probability estimates. *Ann Intern Med.* 2005;143(3):184-189. DOI:10.7326/0003-4819-143-3-20050802 0-00004
3. *Leeflang MM, Deeks J, Gatsonis C, Bossuyt PM; Cochrane*

*Diagnostic Test Accuracy Working Group.* Systematic reviews of diagnostic test accuracy. *Ann Intern Med.* 2008;149(12):889-897. DOI:10.7326/0003-4819-149-12-200812160-00008

4. *Perneger TV, Agoritsas T.* Doctors and patients' susceptibility to framing bias: a randomized trial. *J Gen Intern Med.* 2011;26(12):1411-1417. Doi:10.1007/s11606-011-1810-x

5. *Whiting PF, Davenport C, Jameson C, et al.* How well do health professionals interpret diagnostic information? A systematic review. *BMJ Open.* 2015;5(7):e008155. DOI:10.1136/bmjopen-2015-008155

6. *Parker VL, Ritchie JE, Drake TM, Hookham J, Balasubramanian SP.* A Randomised Assessment of Trainee Doctors' Understanding and Interpretation of Diagnostic Test Results. *World J Surg.* 2016;40(1):21-28. DOI:10.1007/s00268-015-3214-2

7. *Зулькарнаев А.Б.* "Подводные камни" статистического анализа и клинической интерпретации полученных оценок на примере пациентов с хронической болезнью почек. Часть I: оценка риска. *Нефрология и диализ.* 2019. 21(4): 419-429.

*Zulkarnaev A.B.* Pitfalls of statistical analysis and clinical interpretation of the estimates in patients with chronic kidney disease. Part I: risk assessment. *Nefrologiya i dializ [Nephrology and dialysis].* 2019. 21(4): 419-429. DOI: 10.28996/2618-9801-2019-4-419-429

8. *Albarbi FJ, Baig S, Rambhatla SB, et al.* The clinical utility of total concentration of urinary globotriaosylsphingosine plus its analogues in the diagnosis of Fabry disease. *Clin Chim Acta.* 2020;500:120-127. DOI:10.1016/j.cca.2019.10.005

9. *Zhang S, Wu Z, Li J, et al.* Evaluation of the clinical relevance of anti-annexin-A5 antibodies in Chinese patients with antiphospholipid syndrome. *Clin Rheumatol.* 2017;36(2):407-412. DOI:10.1007/s10067-016-3510-8

10. *Levy MM, Fink MP, Marshall JC, Abraham E, Angus D, Cook D, et al.* 2001 SCCM/ESICM/ACCP/ATS/SIS international sepsis definitions conference. *Crit Care Med* 2003;31(4):1250-6. DOI: 10.1007/s00134-003-1662-x

11. *Shapiro NI, Trzeciak S, Hollander JE, et al.* The diagnostic accuracy of plasma neutrophil gelatinase-associated lipocalin in the prediction of acute kidney injury in emergency department patients with suspected sepsis. *Ann Emerg Med.* 2010;56(1):52-59. e1. DOI:10.1016/j.annemergmed.2010.02.010

12. *Рубанович А.В., Хромов-Борисов Н.Н.* Теоретический анализ показателей предсказательной эффективности бинарных генетических тестов. *Экологическая генетика.* 2013; 11(1): 77-90.

*Rubanovich A.V., Khromov-Borisov N.N.* Theoretical analysis of the predictability indices of the binary genetic tests. *Russian Journal of Genetics: Applied Research.* 2013; 11(1): 77-90.

13. *Holmer M, Martínez JP, Gil E, Sandberg F, Olde B, Sörnmo L.* Detection of ventricular premature beats based on the pressure signals of a hemodialysis machine. *Med Eng Phys.* 2018;51:49-55. DOI:10.1016/j.medengphy.2017.11.004

14. *Luo JC, Qiu XH, Pan C, et al.* Increased cardiac index attenuates septic acute kidney injury: a prospective observational study. *BMC Anesthesiol.* 2015;15:22. DOI:10.1186/s12871-015-0005-0

15. *Chang CH, Fu CM, Yang CH, et al.* Society of Thoracic Surgeons score predicts kidney injury in patients not undergoing bypass surgery. *Ann Thorac Surg.* 2015;99(1):123-129. DOI:10.1016/j.athoracsur.2014.07.072

16. *Kim S, Kim HJ, Ahn HS, et al.* Is plasma neutrophil gelatinase-associated lipocalin a predictive biomarker for acute kidney injury in sepsis patients? A systematic review and meta-analysis. *J Crit Care.* 2016;33:213-223. DOI:10.1016/j.jcrc.2016.02.014

17. *Leeftang MM, Deeks JJ, Rutjes AW, Reitsma JB, Bossuyt PM.* Bivariate meta-analysis of predictive values of diagnostic tests can be an alternative to bivariate meta-analysis of sensitivity and specificity. *J Clin Epidemiol.* 2012;65(10):1088-1097. doi:10.1016/j.jclinepi.2012.03.006

Дата получения статьи: 15.09.2020

Дата принятия к печати: 21.11.2020

Submitted: 15.09.2020

Accepted: 21.11.2020