

МЕТОДОЛОГИЯ ОЦЕНКИ ИНФОРМАЦИОННОЙ ЗНАЧИМОСТИ ДИАГНОСТИЧЕСКИХ ТЕСТОВ В АНЕСТЕЗИОЛОГИИ И РЕАНИМАТОЛОГИИ

А. В. Бояркина, А. Л. Потапов

EVALUATION TECHNIQUE FOR INFORMATIVE VALUE OF DIAGNOSTIC TESTS IN ANESTHESIOLOGY AND INTENSIVE CARE

A. V. Boyarkina, A. L. Potapov

Медицинская академия им. С. И. Георгиевского ФГАОУ ВО «Крымский федеральный университет им. В. И. Вернадского», г. Симферополь

S. I. Georgievsky Medical Academy, Simferopol, V. I. Vernadsky Crimea Federal University, Simferopol, Crimea

Выполнен анализ статистических показателей, характеризующих точность диагностических методов и их комбинаций, включая чувствительность, специфичность, позитивную и негативную прогностическую ценность, отношение правдоподобия. Рассмотрен лучший на сегодняшний день метод оценки точности диагностических тестов – построение ROC-кривой и оценка площади под ней (Area Under Curve). Показаны преимущества и недостатки каждого из методов, описаны условия их использования в клинической практике.

Ключевые слова: диагностический тест, комбинированный тест, чувствительность, специфичность, прогностическая значимость, ROC-кривая, площадь под ROC-кривой.

The statistic rates describing the accuracy of diagnostic techniques and their combination have been analyzed including sensitivity, specificity, positive and negative prognostic value, plausibility relation. The best evaluation technique for diagnostic test accuracy at present has been studied, i.e. building up ROC-curve and evaluation of the area under it (Area Under Curve). Advantages and deficiencies of each test have been presented, and their use in the clinical practice has been described.

Key words: diagnostic value, combined test, sensitivity, specificity, prognostic value, ROC-curve, area under curve.

Оценка точности диагностических тестов и правильная интерпретация их результатов в анестезиологии и реаниматологии являются ключевым моментом при постановке диагноза и выборе метода терапии. И врачу, и исследователю важно знать, надо ли проводить тест, какой из них даст более точный ответ, какова вероятность, что у пациента имеется искомое заболевание, изменится ли тактика лечения после получения результатов исследования. Правильная оценка прогностической ценности теста имеет ещё и экономическое значение, так как позволит обосновать затраты на практическое внедрение того или иного метода обследования.

Цель работы – описание некоторых базовых понятий, характеризующих точность диагностических методов, дающих возможность сравнивать тесты между собой и делать выводы о целесообразности проведения той или иной диагностической процедуры.

Чувствительность и специфичность. Существует два базовых понятия, характеризующих точ-

ность диагностических тестов: чувствительность и специфичность. Оба являются одинаково важными и не должны анализироваться по отдельности. Чувствительность (Sensitivity, SE) – это вероятность позитивного результата теста у человека с заболеванием, специфичность (Specificity, SP) – это вероятность негативного результата при отсутствии заболевания у пациента. В данном контексте заболеванием может быть любое состояние или признак, который необходимо диагностировать.

В табл. 1 приведены показатели, необходимые для определения чувствительности и специфичности: TP (True-positive) – это пациенты с заболеванием и положительным результатом теста; FN (False-negative) – пациенты с заболеванием и отрицательным результатом теста; FP (False-positive) – пациенты без заболевания, но с положительным результатом теста; TN (True-negative) – без заболевания с отрицательным результатом теста (табл. 1).

SE и SP будут рассчитываться по следующим формулам [3]:

Таблица 1
Показатели, необходимые для определения SE и SP

Тест	Заболевание/состояние/признак	
	есть	нет
+	Истинно-положительный (TP)	Ложноположительный (FP)
-	Ложноотрицательный (FN)	Истинно-отрицательный (TN)

$$SE = [TP / (TP + FN)] \times 100\%;$$

$$SP = [TN / (TN + FP)] \times 100\%.$$

Исходя из этого, чувствительность – это вероятность TP среди всех пациентов с заболеванием, а специфичность – вероятность TN среди всех пациентов без заболевания. Чем ближе SE и SP к 100%, тем более точным является диагностический тест.

Например, надо оценить SE и SP шкалы тревожности HADS (Hospital Anxiety Depression Scale) для выявления пациентов, склонных испытывать сильную боль после операции (≥ 6 баллов по ЦРШ). Результат ≤ 7 баллов по шкале HADS считался отрицательным, ≥ 8 баллов (наличие симптомов тревоги) – положительным. Из 100 пациентов у 36 результаты теста были истинно-положительными (TP), у 26 пациентов – ложноотрицательными (FN), так как, несмотря на HADS ≤ 7 баллов, интенсивность боли всё равно была ≥ 6 баллов. Следовательно, $SE = 36 / (36 + 26) \times 100\% = 58\%$. У 32 пациентов тест оказался истинно-отрицательным (TN) – при норме по шкале HADS боль была < 6 баллов, однако у 6 пациентов при наличии тревоги боль была низкой интенсивности (FP), следовательно, $SP = 32 / (6 + 32) \times 100\% = 84\%$ (табл. 2).

Совсем немногие тесты одновременно высокочувствительны и высокоспецифичны. Поэтому в ряде случаев пациента подвергают дополнительным обследованиям. В этой ситуации диагностические тесты могут выполнять согласно двум алгоритмам: параллельно (интерпретируются вместе) или последовательно (результаты первого теста определяют, надо ли делать следующий). Преимуществами последовательного метода являются экономия и уклонение от ненужных тестов, а недостатком – возможное удлинение периода обследования, что может негативно повлиять на лечение пациентов в тяжёлом состоянии.

Для интерпретации параллельного алгоритма используют два правила: AND rule и OR rule.

Первое правило исходит из следующего принципа: комбинированный тест считается положительным, если только оба теста (например, тест A и тест B) положительные, а отрицательным – в случае, если отрицательным является хотя бы один из них. Второе правило исходит из другого принципа: если один или оба теста положительные, комбинированный тест считается положительным, комбинированный тест является отрицательным, если только оба теста, A и B, отрицательные [8,10].

При применении правила AND rule расчёт чувствительности и специфичности комбинированного теста (SE_c и SP_c) проводится по формулам:

$$SE_c = SE_A \times SE_B \times 100\%,$$

$$SP_c = [SP_A + SP_B - (SP_A \times SP_B)] \times 100\%,$$

где SE_A , SE_B , SP_A и SP_B – чувствительность и специфичность тестов A и B.

При применении правила OR rule используют формулы:

$$SE_c = [SE_A + SE_B - (SE_A \times SE_B)] \times 100\%,$$

$$SP_c = SP_A \times SP_B \times 100\%.$$

Для примера приведем SE_c и SP_c комбинации шкалы HADS и генотипирования по гену катехол-О-метитрансферазы COMT1947G>A для выявления пациентов с сильной болью после операции при условии, что чувствительность и специфичность генетического теста составляют 88,7 и 36,8% соответственно [1]. При использовании правила AND rule для такой комбинации тестов искомые показатели будут равны: $SE_c = (0,58 \times 0,887) \times 100\% = 51,4\%$, $SP_c = [0,84 + 0,368 - (0,84 \times 0,368)] \times 100\% = 89,9\%$. При применении OR rule – $SE_c = [0,58 + 0,887 - (0,58 \times 0,887)] \times 100\% = 95,3\%$; $SP_c = (0,84 \times 0,368) \times 100\% = 30,9\%$. Таким образом, специалист может заранее сделать выбор в пользу увеличения чувствительности, применяя правило OR rule, и, наоборот, в пользу повышения специфичности, используя правило AND rule, в зависимости от сложившейся клинической ситуации.

Применение высокочувствительных методов целесообразно, когда требуется сузить круг предполагаемых диагнозов. Подобные методики снижают вероятность «пропуска» пациентов, однако дают много «ложных тревог». Высокоспецифичные тесты повышают уверенность, что у пациента не выявлено «несуществующее» заболевание. На наш взгляд, для проведения скрининга пациентов на предмет развития сильной боли после операции более гуманными являются применение высокочувстви-

Таблица 2
Показатели, необходимые для расчета SE и SP шкалы HADS относительно сильной боли после операции

Тест (шкала HADS)	Признак	
	есть (боль ≥ 6 баллов)	нет (боль < 6 баллов)
+ (≥ 8 баллов)	36 (TP)	6 (FP)
- ($0-7$ баллов)	26 (FN)	32 (TN)

тельного теста и использование правила OR rule. При этом максимальное количество пациентов с возможным развитием сильного болевого синдрома будет под пристальным наблюдением специалистов, а число «пропусков» пациентов группы риска – минимальным.

ROC-кривые. В то время как некоторые тесты дают однозначные результаты (положительный или отрицательный), другие показывают различные числовые значения. Например, по вышеупомянутой шкале HADS пациент может набрать от 0 до 21 балла. Для более точной оценки диагностической способности такого теста на начальном этапе необходимо рассмотреть все значения в качестве «порогового» и рассчитать SE и SP для каждой точки. Тест считается положительным при значении HADS выше или равном пороговому значению, отрицательным – при HADS ниже порогового значения. Например, для точки 8 баллов тест считается положительным при $HADS \geq 8$ баллов, а отрицательным – при $HADS < 8$ баллов, для точки 11 баллов тест считается положительным при $HADS \geq 11$ баллов и отрицательным – при $HADS < 11$ баллов. При этом, помимо показателей SE и SP, рассчитывается частота ложноположительных результатов (FPR):

$$FPR = 100 - SP.$$

В табл. 3 приведены значения SE, SP и FPR для нескольких значений по шкале HADS. Данные расчеты можно выполнять вручную или с помощью специальных статистических программ, например MedCalc 15.2 (табл. 3).

На следующем этапе необходимо отобразить полученные значения на графике, результатом чего и является ROC-кривая: по оси X отмечается частота ложноположительных результатов (FPR), а по оси Y – чувствительность (рис. 1). Если ROC-кривая проходит возле диагональной линии, то диагностическая способность теста очень низка

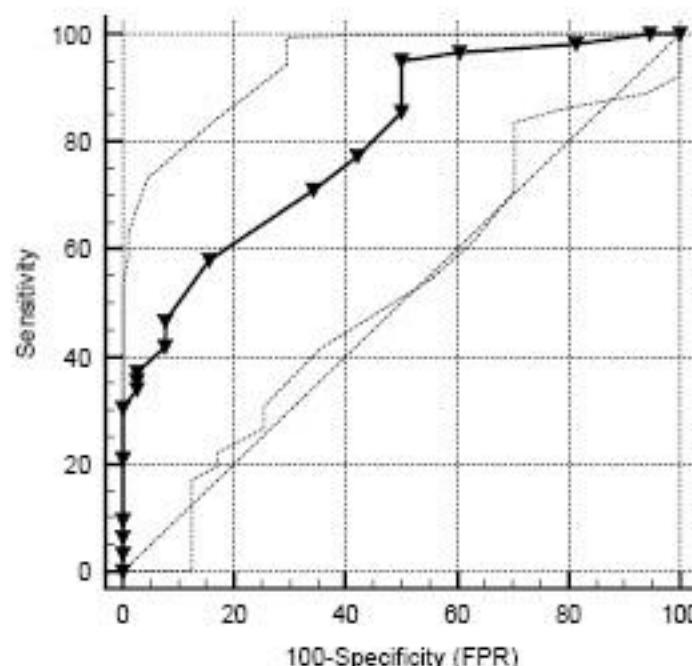


Рис. 1. ROC-кривая шкалы HADS относительно развития сильной боли после операции

и разграничить пациентов с болезнью и без неё будет невозможно. Чем ближе ROC-кривая к левому верхнему углу, тем лучше тест – SE и SP приближаются к 100%, FPR – к 0 [2, 5, 6, 11].

Построенная кривая для шкалы HADS тревоги относительно развития сильной боли после операции занимает среднее положение, что свидетельствует о приемлемых диагностических свойствах.

Для сравнения тестов между собой можно построить одновременно несколько графиков и посмотреть, какой в большей мере стремится к левому верхнему углу (рис. 2).

Визуальное сравнение ROC-кривых не всегда позволяет выявить наиболее эффективную модель. Более точным методом сравнения является оценка площади под ROC-кривыми. Теоретически она изменяется от 0 до 1,0, но поскольку модель всегда характеризуются кривой, расположенной выше положительной диагонали, то обычно говорят об из-

Таблица 3
Показатели SE, SP и FPR для отдельных значений шкалы HADS при выявлении пациентов с сильной болью после операции

HADS (баллы)	SE (%)	SP (%)	FPR
1	98,39	18,42	81,58
3	95,16	50,00	50
5	77,42	57,89	42,11
7	58,06	84,21	15,79
8	46,77	92,11	7,89
10	37,10	97,37	2,63
12	33,87	97,37	2,63
14	20,97	100,00	0
16	6,45	100,00	0
18	3,23	100,00	0

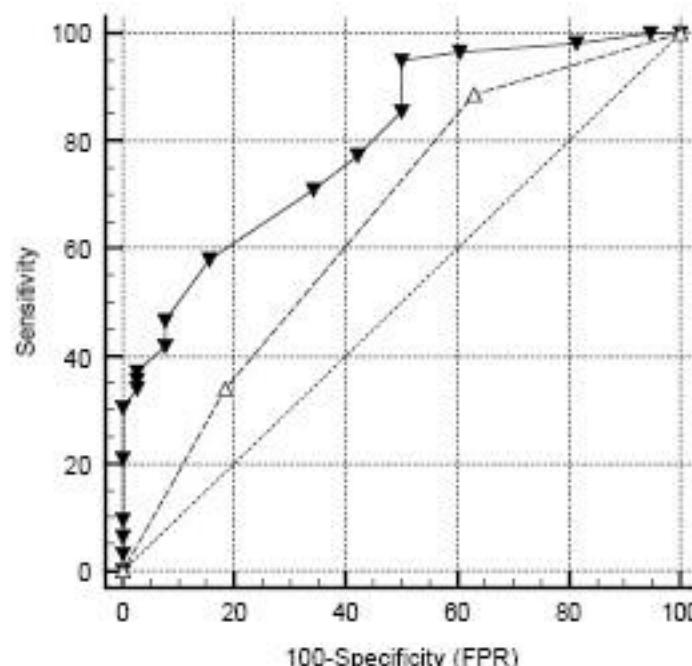


Рис. 2. ROC-кривые шкалы HADS (верхняя) и генотипирования COMT1947G > A (нижняя) относительно развития сильной боли после операции

менениях от 0,5 («бесполезный» классификатор) до 1,0 («идеальная» модель) [6].

Показатель площади под ROC-кривой называется AUC (Area Under Curve). Рассчитать его можно с помощью численного метода трапеций [11] или используя специальные статистические программы. Для оценки полученных значений AUC применяется экспертная шкала, по которой можно судить о качестве модели (табл. 4) [2].

Для шкалы HADS AUC составила 0,802 (95%-ный ДИ 0,710–0,875). Для генотипирования по гену COMT 1947G > A AUC = 0,653 (95%-ный

Таблица 4
Качество прогностической модели в зависимости от AUC

Интервал AUC	Качество модели
0,9–1,0	Отличное
0,8–0,9	Очень хорошее
0,7–0,8	Хорошее
0,6–0,7	Среднее
0,5–0,6	Неудовлетворительное

ДИ 0,551–0,745). Следовательно, более диагностически точным тестом для прогноза сильной боли после операции является шкала HADS тревоги.

Метод расчета AUC может быть использован и для оценки комбинированных тестов, однако для этого требуется освоение более сложных статистических моделей, таких как логистическая регрессия, а также применение современных программ статистического анализа (MedCalc, SPSS и др.).

Таким образом, метод построения ROC-кривых и расчёта AUC является одним из лучших для оценки диагностической точности тестов и всё чаще применяется специалистами при проведении клинических исследований.

Позитивная и негативная прогностическая ценность. Чувствительность и специфичность характеризуют диагностическую значимость теста, но они не дают информации, какова вероятность наличия заболевания у конкретного пациента. В этом случае более уместно использовать такие показатели, как позитивная и негативная прогностическая ценность (PPV, Positive Predictive Value; NPV, Negative Predictive Value), которые отвечают на следующие вопросы: «Какая вероятность, что у пациента есть заболевание при положительном результате теста?» и «Какова вероятность, что его нет при отрицательном?». Значение PPV рассчитывается по формуле [4]:

$$\text{PPV} = \text{TP} / (\text{TP} + \text{FP}).$$

Для примера вернемся к данным табл. 2. Для шкалы HADS PPV = 36/(36 + 6) × 100% = 85,7%. Значение PPV отличается от SE теста, поскольку показывает вероятность развития сильного болевого синдрома у конкретного пациента, а именно:

с тревогой по шкале HADS ≥ 8 баллов. SE отражает вероятность того, что тревога по шкале HADS ≥ 8 баллов присутствует у пациентов с сильной послеоперационной болью.

Таким образом, показатель PPV помогает клиницисту определить тактику лечения при положительном результате теста, а SE характеризует качество диагностического теста и помогает выбрать лучший из них.

Значение NPV рассчитывается по формуле [4]:

$$\text{NPV} = \text{TN} / (\text{TN} + \text{FN}).$$

Исходя из данных табл. 2, для шкалы HADS NPV = 32/(32 + 26) × 100% = 55,2%. Следовательно, у пациента без клинически значимой тревоги (HADS < 8 баллов) не разовьётся сильная боль после операции в 55,2% случаев. NPV отличается от показателя SP, который свидетельствует о вероятности отсутствия тревоги среди всех пациентов без сильной послеоперационной боли.

Значения PPV и NPV могут существенно отличаться в зависимости от распространённости заболевания. Чем ниже распространённость, тем ниже показатели прогностической значимости. В этом случае значения PPV и NPV могут рассчитываться с помощью теоремы Байеса, учитывающей распространённость заболевания в популяции (prevalence – prev.). В этом случае формулы меняются [4, 9]:

$$\text{PPV} = [\text{SE} \times \text{prev} / (\text{SE} \times \text{prev} + (1 - \text{SP}) \times (1 - \text{prev}))] \times 100\%,$$

$$\text{NPV} = [(1 - \text{SP}) \times (1 - \text{prev})] / ((1 - \text{SE}) \times \text{prev} + \text{SP} \times (1 - \text{prev})) \times 100\%.$$

Пусть в данном исследовании распространённость сильной боли после операции составила 62% (0,62). Подставив это значение в формулы, получим, что для HADS тревоги PPV = 85,5%, NPV = 55,1%. Однако если взять распространённость сильной боли в течение первых суток после операции, как в исследовании M. Sommer et al., 2008, где она составила 30%, значения PPV и NPV будут резко отличаться и составят 60,8 и 82,3% соответственно [7]. Данный пример свидетельствует о важности учёта распространённости заболевания вне зависимости от того, какой показатель интересует специалиста в большей мере, PPV или NPV.

Отношение правдоподобия. Дополнительным методом оценки диагностического теста является отношение правдоподобия (LR, Likelihood Ratio). Данное отношение может вычисляться для положительного результата теста (LR+), тогда оно показывает отношение вероятности получить положительный результат у пациента с заболеванием к вероятности получить положительный результат у пациента без него. Отношение правдоподобия для отрицательного результата теста (LR-) показывает отношение вероятности получить отрицательный результат у пациента с заболеванием к вероятности получить отрицательный результат у пациента без него.

Значение LR+ и LR- рассчитываются по формулам [4]:

$$\text{LR}+ = \text{SE}/(1 - \text{SP}),$$

$$\text{LR}- = (1 - \text{SE})/\text{SP}.$$

Значение $\text{LR}+ > 1$ свидетельствует о связи теста с заболеванием, $\text{LR}- < 1$ – о связи теста с отсутствием заболевания.

Для шкалы HADS относительно развития сильной боли (табл. 2) $\text{LR}+ = 0,58/(1 - 0,84) = 3,6$; $\text{LR}- = (1 - 0,58)/0,84 = 0,5$. Полученное для $\text{LR}+$ значение 3,6 свидетельствует о том, что вероятность положительного результата теста ($\text{HADS} \geq 8$ баллов) в 3,6 раза больше у пациента с сильной болью после операции, чем у пациента без болевого синдрома. Значение $\text{LR}-$, равное 0,5, свидетельствует о том, что вероятность получить отрицательный тест ($\text{HADS} < 8$ баллов) у пациента без боли в 2 раза превышает вероятность получить отрицательный тест у пациента с сильной болью после операции.

Таким образом, в настоящее время существует достаточно большой арсенал методов клинической оценки диагностических тестов, начиная от простых, таких как чувствительность, специфичность, позитивная и негативная прогностическая зна-

чимость, и, кончая такими «продвинутыми», как ROC-кривые и AUC. В случае применения комбинированных тестов использование правил OR rule и AND rule позволяет повысить чувствительность или специфичность в зависимости от конкретной клинической ситуации.

ДЛЯ КОРРЕСПОНДЕНЦИИ:

*Медицинская академия им. С. И. Георгиевского
ФГАОУ ВО «Крымский федеральный университет
им. В. И. Вернадского».
295006, Республика Крым, г. Симферополь,
бульвар Ленина, д. 5/7.*

*Бояркина Анна Викторовна
ассистент кафедры анестезиологии-реаниматологии
и скорой медицинской помощи.
E-mail: annaboiarkina@mail.ru*

*Потапов Александр Леонидович
доктор медицинских наук, профессор кафедры
анестезиологии-реаниматологии и скорой медицинской
помощи.
E-mail: ALP8@yandex.ru*

Литература

1. Бояркина А. В. Влияние психологических и генетических факторов на эффективность послеперационной аналгезии опиоидами: Дис. ... канд. мед. наук: 14.01.20. – СПб., 2015. – 137 с. <http://szgmu.ru/ds/upload/files/Диссертация Бояркина.pdf>
2. Паклин Н. Логистическая регрессия и ROC-анализ – математический аппарат [Электрон. ресурс]. – 2015. – Режим доступа: <http://basegroup.ru/community/articles/logistic>
3. Altman D. G., Bland J. M. Diagnostic tests 1: sensitivity and specificity // BMJ. – 1994. – Vol. 308 (6943). – P. 1552.
4. Altman D. G., Bland J. M. Diagnostic tests 2: predictive values // BMJ. – 1994. – Vol. 309 (6947). – P. 102.
5. DeLong E. R., DeLong D. M., Clarke-Pearson D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach // Biometrics. – 1988. – Vol. 44, № 3. – P. 837–845.
6. Hanley J. A., McNeil B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve // Radiology. – 1982. – Vol. 143, № 1. – P. 29–36.
7. Sommer M., de Rijke J. M., van Kleef M. et al. The prevalence of postoperative pain in a sample of 1490 surgical inpatients // Eur. J. Anaesthesiol. – 2008. – Vol. 25, № 4. – P. 267–274.
8. Parikh R., Mathai A., Parikh S. et al. Understanding and using sensitivity, specificity and predictive values // Indian. J. Ophthalmol. – 2008. – Vol. 56, № 1. – P. 45–50.
9. Weinstein S., Obuchowski N. A., Lieber M. L. Clinical evaluation of diagnostic tests // Am. J. Roentgenol. – 2005. – Vol. 184, № 1. – P. 14–19.
10. Zhou X. H., Obuchowski N. A., McClish D. K. Statistical methods in diagnostic medicine // Biometrics. – 2003. – Vol. 59, № 1. – P. 203–204.
11. Zweig M. H., Campbell G. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine // Clin. Chem. – 1993. – Vol. 39, № 4. – P. 561–577.

References

1. Boyarkina A.V. Vliyanie psichologicheskikh i geneticheskikh faktorov na effektivnost' posleoperatsionnoy analgezii opioidami. Diss. kand. med. nauk. [Effect of psychological and genetic factors on the efficiency of post-surgery analgesia with opioids. Cand. Diss.]. 14.01.20. St. Petersburg, 2015, 137 p. <http://szgmu.ru/ds/upload/files/Диссертация Бояркина.pdf>
2. Paklin N. Logisticheskaya regressiya i ROC-analiz – matematicheskiy apparat. [Logic regression and ROC-analysis – mathematical tool]. (Epub.) 2015, Available at <http://basegroup.ru/community/articles/logistic>
3. Altman D.G., Bland J.M. Diagnostic tests 1: sensitivity and specificity. *BMJ*, 1994, vol. 308 (6943), pp. 1552.
4. Altman D.G., Bland J.M. Diagnostic tests 2: predictive values. *BMJ*, 1994, vol. 309 (6947), pp. 102.
5. DeLong E.R., DeLong D.M., Clarke-Pearson D.L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 1988, vol. 44, no. 3, pp. 837–845.
6. Hanley J.A., McNeil B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 1982, vol. 143, no. 1, pp. 29–36.
7. Sommer M., de Rijke J.M., van Kleef M. et al. The prevalence of postoperative pain in a sample of 1490 surgical inpatients. *Eur. J. Anaesthesiol.*, 2008, vol. 25, no. 4, pp. 267–274.
8. Parikh R., Mathai A., Parikh S. et al. Understanding and using sensitivity, specificity and predictive values. *Indian. J. Ophthalmol.*, 2008, vol. 56, no. 1, pp. 45–50.
9. Weinstein S., Obuchowski N.A., Lieber M.L. Clinical evaluation of diagnostic tests. *Am. J. Roentgenol.*, 2005, vol. 184, no. 1, pp. 14–19.
10. Zhou X.H., Obuchowski N.A., McClish D.K. Statistical methods in diagnostic medicine. *Biometrics*, 2003, vol. 59, no. 1, pp. 203–204.
11. Zweig M.H., Campbell G. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clin. Chem.*, 1993, vol. 39, no. 4, pp. 561–577.