



О.К. ПОТАНИНА,

врач НИИ СП им. Н.В. Склифосовского, г. Москва, olga-honta@rambler.ru

А.Г. ДОРФМАН,

д.м.н., профессор, главный научный сотрудник НИИ СП им. Н.В. Склифосовского, г. Москва, dorfman@mail.ru

Е.В. ОГУРЦОВА,

аспирант Института Макса Планка, г. Росток, Германия, kogurtsova@yahoo.com

С.Л. ШВЫРЕВ,

к.м.н., старший преподаватель кафедры медицинской кибернетики и информатики
ГОУ ВПО РГМУ Росздрава, г. Москва, interis@bk.ru

Т.В. ЗАРУБИНА,

д.м.н., профессор, заведующая кафедрой медицинской кибернетики и информатики
ГОУ ВПО РГМУ Росздрава, г. Москва, t_zarubina@mail.ru

СРАВНЕНИЕ ЭФФЕКТИВНОСТИ ПРОГНОСТИЧЕСКИХ ШКАЛ ОЦЕНКИ ТЯЖЕСТИ СОСТОЯНИЯ РЕАНИМАЦИОННЫХ БОЛЬНЫХ ХИРУРГИЧЕСКОГО ПРОФИЛЯ

УДК 616-036

Потанина О.К., Дорфман А.Г., Огурцова Е.В., Швырев С.Л., Зарубина Т.В. Сравнение эффективности прогностических шкал оценки тяжести состояния реанимационных больных хирургического профиля (Государственное образовательное учреждение высшего профессионального образования «Российский государственный медицинский университет Федерального агентства по здравоохранению и социальному развитию»; Департамент здравоохранения г. Москвы; Государственное учреждение здравоохранения г. Москвы «Научно-исследовательский институт скорой помощи имени Н.В. Склифосовского»)

Аннотация: Целью работы явилась оценка и сравнение эффективности прогнозирования летального исхода пяти шкал для оценки тяжести состояния (APACHE II, APACHE IV, SAPS II, MPM II 0 и POSSUM) для группы реанимационных больных хирургического профиля в условиях российского стационара. В основу анализа легло проспективное исследование, проведенное на базе отделения реанимации и интенсивной терапии (ОРИТ) для послеоперационных больных НИИ СП им. Н.В. Склифосовского. Были собраны клинические и физиологические данные 300 пациентов на момент поступления и по истечению 1-х суток пребывания пациента в ОРИТ, необходимые для расчета прогностических моделей. В процессе анализа сравнивались дискриминационная способность шкал с помощью построения ROC-кривых, а также их калибрационная способность с использованием теста Хосмера-Лемешоу и построения калибровочных кривых. Анализ ROC кривых показал, что шкалы APACHE II, APACHE IV, SAPS II, MPM II 0 имеют хорошую дискриминационную способность (площадь под ROC-кривой > 0,9), а шкала POSSUM — удовлетворительную дискриминационную способность (площадь под ROC-кривой > 0,8). Удовлетворительная калибрационная способность была выявлена только у двух из пяти прогностических шкал: SAPS II и MPM II. Таким образом, прогностические шкалы SAPS II и MPM II 0, имеющие высокую дискриминационную способность и удовлетворительную калибровку, могут быть рекомендованы для оценки вероятности летального исхода в ОРИТ хирургического профиля.

Ключевые слова: APACHE II; APACHE IV; SAPS II; MPM II 0; POSSUM; прогностические шкалы.

UDC 616-036

Potagina O., Dorfman A., Ogurtsova E., Shvirev S., Zarubina T. Estimation efficiency and comparison of several prognostic scoring systems for ICU patients of surgical profile (The Russian State Medical University; N. V. Sklifosovsky Research Institute of Emergency Medicine)

Abstract: The aim of the study was the estimation efficiency and the comparison several prognostic scoring systems for severity conditions estimation (APACHE II, APACHE IV, SAPS II, MPM II 0 and POSSUM) for ICU patients of surgical profile in a Russian hospital. The analysis was constructed on the prospective study that was conducted on the Intensive Care Unit for Postoperative patients of N. V. Sklifosovsky Research Institute of Emergency Medicine. Clinical and physiological data required for the calculation of predictive models were collected on the status of 300 patients on admission to ICU and during the first 24 hours of sojourn time in the department. The discriminatory ability of models were compared by constructing of ROC-curves, as well as their calibration were analyzed by applying Hosmer-Lemeshow's test and the construction of calibration curves. The results showed scales of APACHE II, APACHE IV and POSSUM have poor calibration, SAPS II and MPM II 0 have satisfactory calibration. Analysis of ROC curves showed that the scale of APACHE II, APACHE IV, SAPS II, MPM II 0 have a good discriminatory ability (area under the ROC-curve > 0,9), and the scale of POSSUM satisfactory discriminatory ability (area under the ROC-curve > 0,8). At the same time we didn't find significant differences in discriminatory ability between all scales by applying the DeLong's test to compare the areas under the ROC curves for different models. Thus, according our results the scales SAPS II and MPM II 0 have the satisfactory calibration and high discriminatory power and can be recommended to estimate the probability of death in ICU of surgical profile.

Keywords: APACHE II; APACHE IV; SAPS II; MPM II 0; POSSUM; prognostic scoring systems.



Применение информационных технологий в повседневной практике отделений реанимации и интенсивной терапии (ОРИТ) давно стало общепринятым. В настоящее время в ОРИТ все чаще используются системы, которые не только обеспечивают интенсивное наблюдение за клиническими и лабораторными данными пациента, но и оказывают поддержку в процессе интерпретации данных и принятия врачебных решений. Среди первоочередных задач, которые постоянно приходится решать врачу при ведении реанимационных больных, можно выделить оценку тяжести состояния.

Оценка тяжести состояния пациента — принципиально важный момент в работе врача-реаниматолога, необходимый при определении тактики ведения пациента ОРИТ. Современные прогностические шкалы позволяют получить объективизированную оценку тяжести состояния пациента на основе расчета вероятности летального исхода и вероятности послеоперационных осложнений, способны определить ориентировочные сроки пребывания пациента в ОРИТ и некоторые другие параметры [1, 8].

Прогностические шкалы используют не только для объективизированной оценки тяжести состояния конкретного пациента, но и для анализа эффективности работы ОРИТ при сравнении результатов работы различных ОРИТ, сопоставлении эффективности различных видов терапии и методов лечения. Публикуемые научные материалы и результаты исследований, при которых необходима рандомизация групп пациентов для получения достоверных сопоставимых данных, также базируются на результатах расчета прогностических шкал [2, 3].

За последние тридцать лет создано несколько десятков шкал оценки тяжести состояния пациентов, большая часть которых используется в практике ОРИТ. Среди них существуют как универсальные модели расчета вероятности летального исхода, (APACHE, SAPS, MPM, ODIN) [8, 9, 10, 11], так и модели, ориентированные на оценку тяжести состояния больных при некоторых заболеваниях: EUROSCORE, ONTARIO в кардиохирургии [12, 13], TRISS, ISS, RTS в травматологии [1], MPM для онкологических больных и другие [14].

Большинство прогностических шкал разработаны на основе логистического мультивариантного анализа и отличаются друг от друга набором клинических, инструментальных и лабораторных параметров и особенностями их использования. Количество и степень доступности этих параметров, используемых в качестве входных переменных, во многом определяет удобство и качество использования прогностических шкал в повседневной клинической практике.

В зарубежной литературе существует большое количество публикаций, посвященных использованию прогностических шкал для оценки тяжести состояния реанимационных больных [1]. Результаты многочисленных сравнительных исследований прогностической точности шкал, проведенных во многих странах на разных категориях пациентов, носят противоречивый характер, что объясняется неодинаковым удельным весом прогностических факторов в различных стационарах, изменением предсказательной мощности этих факторов во времени и многими другими причинами. В отечественной литературе работы, посвященные сравнительному анализу прогностической эффективности шкал в условиях российских стационаров, представлены недостаточно. По больным хирургического профиля опубликованы только единичные работы [4, 5, 6]. Такое положение дел обусловило проведение сравнительного анализа эффективности и точности существующих прогностических моделей в условиях отечественного ОРИТ.

Материалы и методы

Данное проспективное исследование было проведено в период с июля 2009 по сентябрь 2010 года на базе отделения реанимации и интенсивной терапии для послеоперационных больных НИИ СП им. Н.В. Склифосовского. В проспективную выборку включались все поступившие в ОРИТ и находящиеся в нем более 4 часов пациенты старше 18 лет с хирургической патологией. Из выборки исключались пациенты с ожогами или кардиохирургической патологией, а также пациенты, повторно поступившие в ОРИТ или переведенные в другой стационар.

Всего было обследовано 300 пациентов, средний возраст которых составил $56,3 \pm 20,0$ лет. Лиц





Таблица 1

Характеристика обследованных пациентов

Нозологические группы	Всего больных		Летальность в группе		Количество осложнений в группе	
	Количество	%	Количество	%	Количество	%
Травмы различных локализаций	74	25	4	5	22	30
Желудочно-кишечное кровотечение	39	13	8	21	10	26
Аневризмы брюшного и/или грудного отделов аорты	36	12	22	61	10	28
Эмболии и тромбозы артерий и вен верхних и нижних конечностей	34	11	5	15	6	18
Прочие	117	39	19	16	27	23

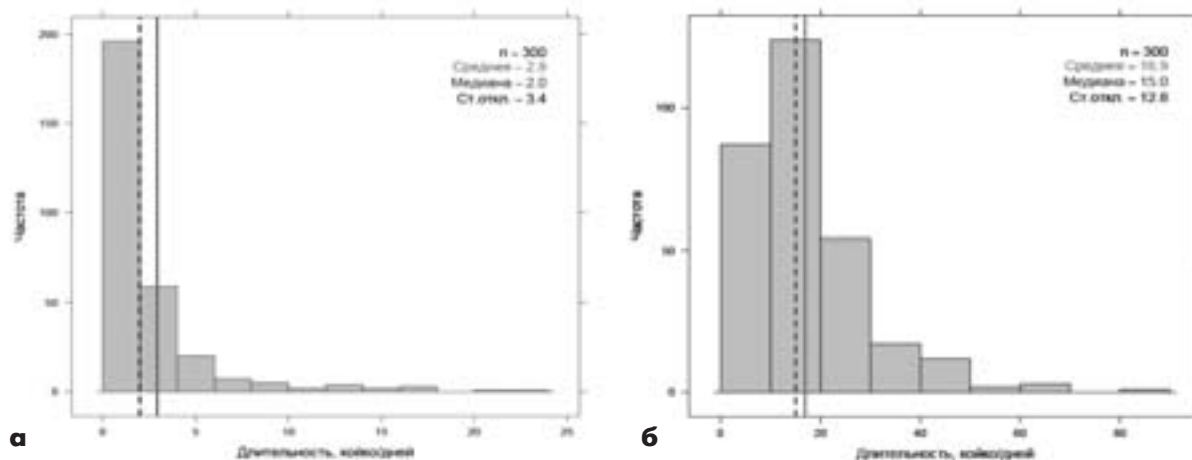


Рис. 1. Распределение пациентов по срокам пребывания в ОРИТ (а) и длительности госпитализации в стационаре (б)

мужского пола было 186 (62%), женского пола — 114 (38%). 170 пациентов (56,7%) перенесли хирургическое вмешательство, 130 пациентам (43,3%) проводилась интенсивная консервативная терапия. Летальность составила 19,3%.

По нозологическим группам пациенты распределились следующим образом: травмы различной локализации — 74 пациента, желудочно-кишечные кровотечения — 39, острая кишечная непроходимость — 17, перфорация кишечника — 13, аневризма брюшного и/или грудного отделов аорты — 36 пациентов, тромбозы или эмболии сосудов, включая массивную тромбоэмболию легочной артерии, — 34 пациента (табл. 1).

У большинства пациентов длительность пребывания в ОРИТ была не более 4–5 койко-дней, а средняя длительность госпитализации составила $16,9 \pm 12,8$ койко-дней (рис. 1). Всем поступившим пациентам рассчитывались шкалы APACHE II, APACHE IV, SAPS II, MPM II 0. Шкала POSSUM, включающая интраоперационные данные, вычислялась у пациентов, поступивших в ОРИТ после хирургических вмешательств. Сбор клинических, лабораторных и инструментальных данных, необходимых для расчета прогностических шкал, производился в течение 24 часов после поступления пациента в ОРИТ в полном соответствии с требованиями разработчиков. Обязательным



Таблица 2

Результаты прогнозирования летального исхода

Шкала	Среднее значение	Ошибка среднего*	Станд. отклонение	Перцентили				
				0%	25%	50%	75%	100%
APACHE II	17,5	2,1	21,9	0,0	4,0	9,0	20,3	96,0
APACHE IV	10,2	2,0	20,7	0,0	1,0	2,0	5,3	99,0
SAPS II	17,1	2,4	25,6	0,0	2,0	6,0	17,3	99,0
MPM II 0	20,3	2,1	22,5	1,0	6,0	11,0	24,0	99,0
POSSUM	24,2	2,2	23,1	2,0	7,0	15,5	32,0	88,0

* ошибка среднего рассчитывалась при уровне достоверности 0,95

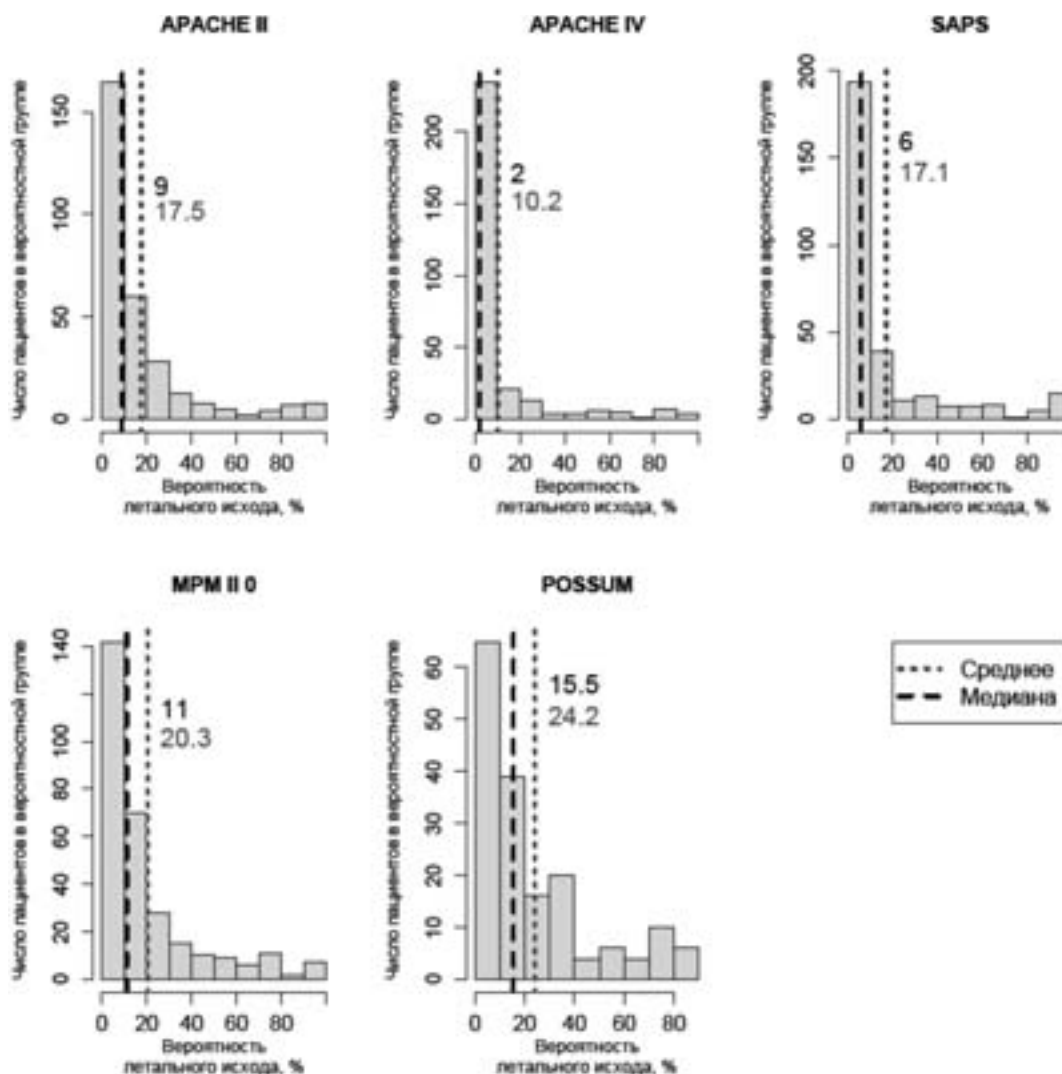


Рис. 2. Результаты расчета вероятности летального исхода с помощью пяти прогностических шкал (столбцы отражают количество пациентов в различных вероятностных группах)





условием расчета всех шкал было наличие полного набора входных параметров. Заполнение пропусков не допускалось. Результаты расчета прогностических шкал не были известны врачам, оказывающим медицинскую помощь пациентам ОРИТ, и не могли повлиять на результаты лечения.

Вычисления проводились с использованием информационной системы отделения реанимации и интенсивной терапии «ИНТЕРИС», которая в течение многих лет успешно эксплуатируется в нескольких ОРИТ [7]. В «ИНТЕРИС» разработан специализированный модуль оценки тяжести состояния реанимационных больных, в котором представлена справочная информация по сбору данных, реализованы и протестированы авторские алгоритмы расчета прогностических шкал, созданы формы сохранения и представления полученных результатов.

В результате расчета всех прогностических шкал вычислялась вероятность летального исхода. При расчете шкалы POSSUM дополнительно вычислялась вероятность развития послеоперационных осложнений. При расчете шкалы APACHE IV дополнительно вычислялась ориентировочная длительность пребывания пациента в ОРИТ.

Калибрационная способность шкал проверялась с использованием С-критерия и Н-критерия согласия Хосмера-Лемешоу. Дискриминационная способность шкал определялась методом построения ROC-кривых и определения порогового значения для оптимального соотношения чувствительности и специфичности. Для оценки достоверности различий дискриминационной способности прогностических шкал использовался критерий ДеЛонга [20]. Для статистического анализа использовалась свободная программная среда для статистического анализа и графического представления данных на основе вычислений с открытым исходным кодом R, version 2.12.1 [17].

Результаты исследования и их обсуждение

Усредненный прогноз летального исхода, определенный с помощью пяти прогностических шкал на выборке, включающей 300 пациентов, оказался сопоставимым с наблюдаемой в этой

группе летальностью. Наиболее близкими оказались результаты, полученные по шкалам APACHE II, SAPS II и MPM II 0 (табл. 2). Прогноз, полученный по шкале APACHE IV, был слишком оптимистичным ($10,2 \pm 2,0\%$). Результат расчета шкалы POSSUM, напротив, несколько превышал наблюдаемую летальность ($24,2 \pm 2,2\%$).

Рассчитанные медианы отражают аналогичную тенденцию: APACHE IV (медиана — 9,0%; процентиль 25 — 4,0%, процентиль 75 — 20,3%), SAPS II (медиана — 6,0%; процентиль 25 — 2,0%, процентиль 75 — 17,3%), MPM II 0 (медиана — 11,0%; процентиль 25 — 6,0%, процентиль 75 — 24,0%). Значительное смещение медианных значений влево объясняется тем, что подавляющее большинство пациентов попадает в первую группу с расчетной вероятностью летального исхода от 0 до 10%. Особенно это заметно в результатах расчета прогностической шкалы APACHE IV (рис. 2).

Вопрос о возможности и целесообразности использования прогностической шкалы в повседневной клинической практике напрямую связан с качеством полученных результатов, определяющих ее прогностическую точность для данной категории пациентов ОРИТ. Качество прогностической шкалы основывается на двух основных характеристиках: калибрационной способности (calibration) и дискриминационной способности (discrimination).

Калибрационная способность отражает способность шкалы предсказывать летальный исход во всех вероятностных группах. Строго говоря, она является проверкой на согласие и подразумевает проверку предположения о том, что наблюдаемая летальность подчиняется тому же закону распределения, который описывается математической моделью, лежащей в основе прогностической шкалы.

Анализ калибрационной способности прогностических шкал проводился с использованием С-критерия и Н-критерия Хосмера-Лемешоу. С-критерий включает в себя разделение пациентов на 10 групп, содержащих равное количество человек в порядке возрастания прогнозируемой вероятности летального исхода. При этом ширина вероятностного интервала варьирует в зависимости от вероятности летального исхода у пациентов, попавших в группу. Для Н-критерия все 10



Таблица 3

Результаты оценки калибрационной способности прогностических шкал с использованием С- и Н-критериев Хосмера-Лемешоу

Шкала	Тест Хосмера-Лемешоу					
	С-критерий			Н-критерий		
	Значение χ^2	Степени свободы	р-значение	Значение χ^2	Степени свободы	р-значение
APACHE II	23,8	8	0,0025*	23,2	8	0,0031*
APACHE IV	45,6	2	0,001*	67,7	8	0,001*
SAPS II	12,7	8	0,1244	14,5	8	0,0695
MPM II 0	9,4	8	0,3094	10,7	8	0,2185
POSSUM	27,3	8	0,0006*	27,2	8	0,0001*

* — при $p < 0,05$ нулевая гипотеза о сходстве распределений отвергается.

групп являются фиксированными по вероятностным интервалам с различным количеством пациентов в группах. Различие между наблюдаемой и ожидаемой летальностью для каждой шкалы оценивается с использованием χ^2 -критерия Пирсона. Статистики по каждой группе суммируются и сравниваются с χ^2 -распределением для степени свободы, которая принимается на две меньше, чем количество вероятностных групп.

Три из пяти моделей (APACHE II, APACHE IV, POSSUM) показали неудовлетворительную калибрационную способность на данной категории пациентов. Рассчитанные значения χ^2 для С- и Н-критериев теста Хосмера-Лемешоу достоверно различаются с уровнем значимости $p < 0,05$, таким образом, нулевая гипотеза о сходстве распределений прогнозируемой и наблюдаемой летальности отвергается. Эти результаты указывают на то, что существует значительная разница между наблюдаемой и ожидаемой летальностью, и, как следствие, на недостаточную калибрационную способность этих прогностических шкал. В то же время для моделей SAPS II и MPM II 0 тест не показал достоверных различий между распределениями прогнозируемой и наблюдаемой летальности (табл. 3).

Калибрационные способности моделей можно оценить визуально с помощью графического представления прогнозируемых и наблюдаемых исходов заболевания в различных вероятностных группах риска летального исхода (рис. 3).

Для построения этих кривых пациенты были разделены на группы, основанные на прогнозируемой вероятности летального исхода (0–10%, 11–20% и т.д.). Реальная летальность в каждой группе (число умерших по отношению к общему количеству пациентов в группе) графически сравнивалась с линией, соответствующей ожидаемой летальности. Для модели с хорошей калибрационной способностью калибровочная кривая должна проходить рядом с этой линией (на графике — диагонали). Четыре из пяти шкал (APACHE II, APACHE IV, SAPS II, MPM II 0) продемонстрировали тенденцию к недооценке тяжести состояния пациентов, то есть прогнозировали летальность, ниже наблюдаемой во всех группах, за исключением первых двух групп с минимальным значением вероятности летального исхода. Шкала POSSUM, наоборот, давала завышенный прогноз летальности во всех вероятностных группах. Визуальный анализ графиков позволяет заключить, что результаты расчета шкалы MPM II 0 наиболее близко расположены к линии ожидаемой летальности (наблюдаемая летальность близка к ожидаемой) по сравнению с другими прогностическими шкалами. Несмотря на то, что прогностические модели показывают результаты, близкие к наблюдаемым значениям летальности в первых двух вероятностных группах, существуют значимые вариации в последующих вероятностных группах, что может быть связано с относительно малым числом пациентов в них.

Таким образом, анализ калибрационной способности изученных нами прогностических шкал



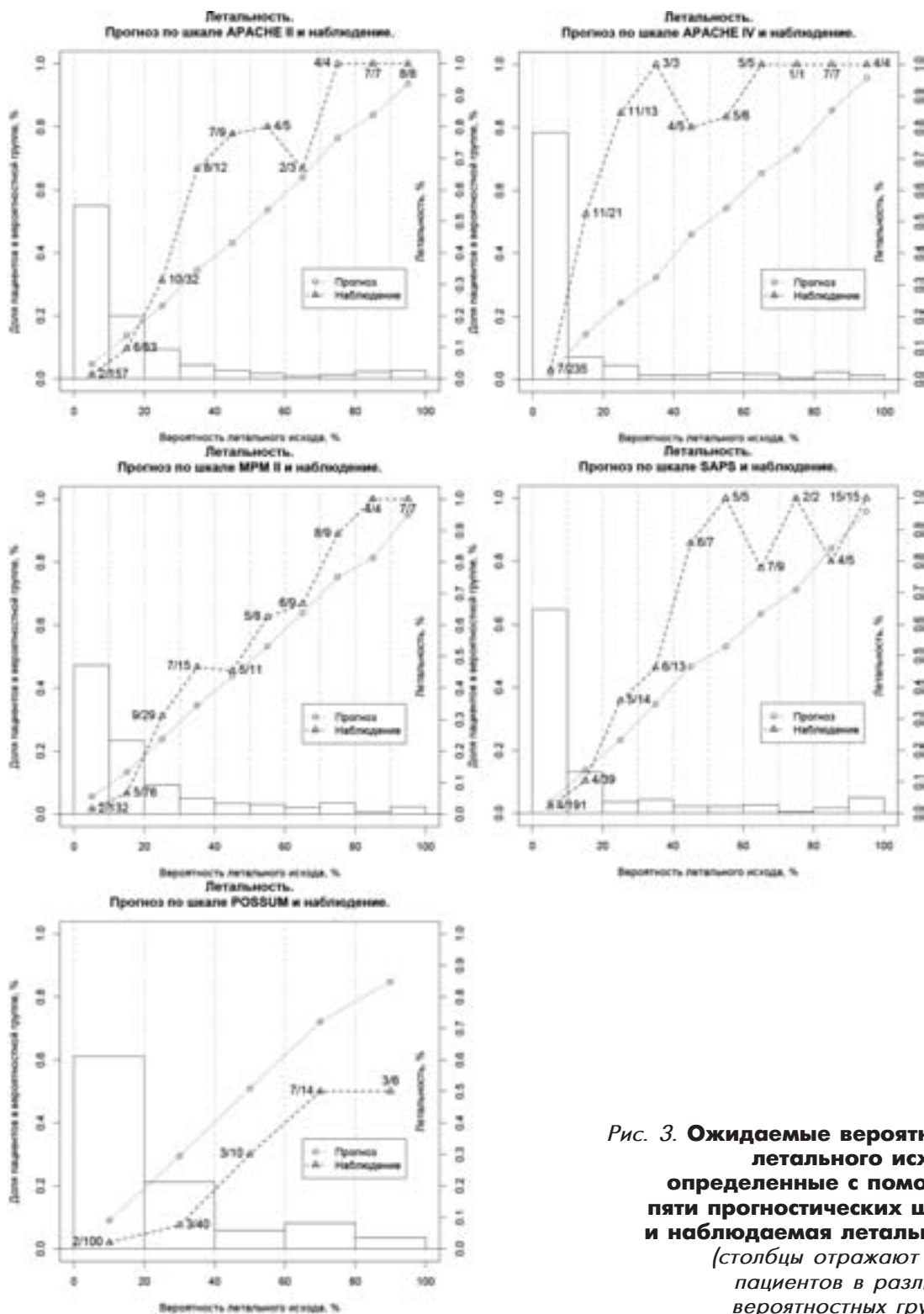


Рис. 3. Ожидаемые вероятности летального исхода, определенные с помощью пяти прогностических шкал, и наблюдаемая летальность (столбцы отражают долю пациентов в различных вероятностных группах)



Таблица 4

Дискриминационный анализ. Площадь под ROC-кривой

Шкала	Площадь под ROC-кривой	р-значение*	95% доверительн. интервал	Стандартная ошибка**
APACHE II	0,941	<0,001	0,904–0,978	0,019
APACHE IV	0,964	<0,001	0,938–0,990	0,013
SAPS II	0,942	<0,001	0,903–0,980	0,019
MPM II 0	0,910	<0,001	0,866–0,954	0,022
POSSUM	0,861	<0,001	0,754–0,968	0,053

* — при $p < 0,05$ нулевая гипотеза о том, что площадь под кривой равна 0.5, отвергается.

** — непараметрическая оценка.

Таблица 5

Результаты сравнения площадей под ROC-кривыми для нескольких прогностических шкал с помощью двустороннего критерия ДеЛонга (Z)

Шкалы	APACHE IV		SAPS II		MPM II 0		POSSUM	
	Z	p	Z	p	Z	p	Z	p
APACHE II	-1,336	0,182	-0,032	0,974	1,665	0,096	0,619	0,536
APACHE IV			1,619	0,106	2,955	0,003*	1,266	0,206
SAPS II					2,756	0,006*	0,146	0,884
MPM II 0							0,208	0,835

* — при $p < 0,05$ нулевая гипотеза о том, что разница между двумя значениями площади под кривой равна 0, отвергается.

позволяет заключить, что только две из них (MPM II 0 и SAPS II) имеют удовлетворительную калибрационную способность для данной группы пациентов ОРИТ, в то время как у других шкал она явно недостаточна.

Разрешающая (дискриминационная) способность диагностического или прогностического теста по делению случаев на две группы (наличие и отсутствие болезни в случае диагностики и умершие/выжившие в случае прогноза) оценивается путем анализа характеристической кривой (Receiver Operating Characteristic — ROC) [15, 16]. ROC-кривая для бинарной классификационной системы отражает зависимость ее чувствительности (доли истинно положительных предсказаний) от специфичности (доли ложноположительных предсказаний). Для количественной оценки разрешающей способности используется площадь под ROC-кривой (AUROC — Area Under Receiver Operator Curve), значения которой находятся в пределах от 0,5 (разрешающая

способность по воле случая, 50 на 50%) до 1,0 (идеальная разрешающая способность). В случае сравнения разрешающей способности прогностических шкал лучшей считается та, у которой площадь под ROC-кривой больше. Дискриминационная способность считается достаточной, если площадь под ROC-кривой выше 0,80 [19]. При значении выше 0,90 разрешающая способность прогностической шкалы оценивается как высокая.

Анализ полученных нами результатов показал, что все шкалы при прогнозировании летального исхода обладают статистически значимой дискриминационной способностью, так как площадь под ROC-кривой для всех моделей была достоверно больше 0,5. Шкала POSSUM обладает достаточной дискриминационной способностью, а остальные шкалы продемонстрировали высокую дискриминационную способность (табл. 4).

Сравнение дискриминационной способности прогностических шкал между собой с использова-





Таблица 6

Основные характеристики прогностических шкал на уровне отсечения 0,5 и оптимальной точке отсечения

Характеристики шкал	APACHE II		APACHE IV		SAPS II		MPM II 0		POSSUM	
Уровень отсечения	0,50	0,15	50	0,04	50	0,12	50	0,15	50	0,18
Чувствительность	0,431	0,948	0,397	0,931	0,569	0,914	0,517	0,914	0,611	0,944
Специфичность	0,992	0,814	0,996	0,843	0,988	0,802	0,971	0,727	0,901	0,599
TP	25	55	23	54	33	53	30	53	11	17
TN	240	197	241	204	239	194	235	176	137	91
FP	2	45	1	38	3	48	7	66	15	61
FN	33	3	35	4	25	5	28	5	7	1
TN + TP	265	252	264	258	272	247	265	229	148	108
SMR	2148	0,58	2417	0,63	1611	0,574	1568	0,487	0,692	0,231
PPV	0,926	0,55	0,958	0,587	0,917	0,525	0,811	0,445	0,423	0,218
NPV	0,879	0,985	0,873	0,981	0,905	0,975	0,894	0,972	0,951	0,989

TN (true negative) — истинно отрицательные результаты (выздоровевшие пациенты получили отрицательный прогноз).

TP (true positive) — истинно положительные результаты (пациенты с заболеванием имеют положительный результат теста)

FN (false negative) — ложнонегативный результат (при наличии заболевания оно не диагностируется) (ошибка второго рода)

FP (false positive) — ложноположительный результат (тест дает положительный ответ при отсутствии заболевания у пациента). (ошибка первого рода)

Se (Sensitivity) — чувствительность, или доля истинно положительных $TPR = TP/P = TP/(TP+FN)$

Sp (Specificity) — специфичность или доля истинно отрицательных $SPC = TN/N = TN/(FP+TN) = 1 - FPR$

SMR (Standardized mortality ratio) — стандартизированное отношение смерти $(TP+FN)/(TP+FP) = PPV/Sens$ — отношение наблюдаемого количества летальных исходов к прогнозируемому числу летальных исходов

PPV (Positive Predictive Value) — ценность положительного прогноза = $TP / (TP + FP)$

NPV (Negative predictive value) — ценность отрицательного прогноза $NPV = TN / (TN + TP)$

нием критерия ДеЛонга не выявило явных лидеров и аутсайдеров: ни одна из них не отличалась статистически достоверно от всех остальных (табл. 5).

При проведении парного сравнения оказалось, что дискриминационная способность шкалы MPM II 0 достоверно ниже дискриминационной способности моделей APACHE IV и SAPS II. В то же время дискриминационные способности шкал MPM II 0 и APACHE II не имеют между собой статистически значимых различий.

С практической точки зрения, наиболее важной является способность прогностической шкалы с высокой степенью точности предсказывать результат лечения каждого поступившего в ОРИТ или, другими словами, строить индивидуальный прогноз пациента. Высокая дискриминационная способность четырех шкал позволяет использовать их для

этой цели. Условное разделение поступающих в ОРИТ на пациентов с потенциально благоприятным и потенциально неблагоприятным результатом лечения производится по величине вероятности летального исхода. Для этого разделения используется пороговое значение вероятности — уровень отсечения (cut-off), при превышении которого пациент может быть условно отнесен к группе с летальным исходом, а в противном случае — к группе выздоравливающих больных.

Определение оптимального уровня отсечения для каждой прогностической шкалы производится с помощью критериев выбора порогового значения вероятности, большинство из которых базируются на основе искомого соотношения чувствительности и специфичности. Использование шкал в ОРИТ в первую очередь нацелено на выявление

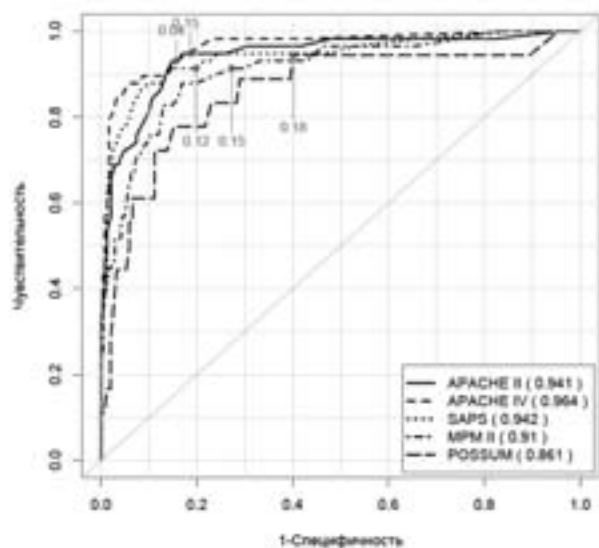


Рис. 4. ROC-кривые, построенные для пяти прогностических шкал, с оптимальным порогом отсечения (в скобках указана площадь под ROC-кривой)

группы риска — пациентов с летальным исходом (положительный прогноз). Чтобы не пропустить таких пациентов, необходимо минимизировать количество ложноотрицательных прогнозов за счет высокой чувствительности прогностического инструмента. В то же время специфичность должна быть также достаточно высокой, так как большое количество пациентов, ошибочно отнесенных к группе с летальным исходом (ложноположительный прогноз), ослабляют бдительность врачей и медицинского персонала ОРИТ.

При уровне отсечения 0,5, который исходно представлялся нам наиболее разумным, все прогностические шкалы продемонстрировали неудовлетворительный уровень чувствительности и очень высокую специфичность (табл. 6). Полученные результаты продиктовали необходимость поиска оптимального порогового значения путем его смещения влево. Критерием выбора оптимального порогового значения было сочетание значений чувствительности выше 0,9 и специфичности не ниже 0,8. Для трех из пяти шкал уровни отсечения в соответствии с этим критерием были найдены: APACHE II (0,15), APACHE IV (0,04), SAPS II (0,12). Для двух других шкал (MPM II 0 и POSSUM)

высокий уровень чувствительности обеспечивался за счет снижения специфичности (рис. 4).

Заключение

Одной из важных задач, которую приходится решать врачам ОРИТ, является комплексная оценка состояния пациента на основе клинических наблюдений, лабораторных данных и результатов инструментальных методов исследования. Современные информационные системы, обеспечивающие поддержку принятия врачебных решений при лечении пациентов в ОРИТ, используют объективизированные шкалы оценки тяжести состояния пациента. В настоящее время создано большое количество шкал оценки тяжести состояния пациента, и очевидной является необходимость проведения исследований их валидности в условиях Российской Федерации.

Результаты проведенного нами исследования показали, что все прогностические модели, за исключением шкалы POSSUM, имеют тенденцию к недооценке наблюдаемой летальности. Особенно это заметно в вероятностных группах с прогнозируемой вероятностью летального исхода выше 0,2, в которые попадает относительно небольшое число пациентов и отмечается наибольшее расхождение между прогнозируемой и наблюдаемой летальностью.

Шкалы MPM II 0 и SAPS II показали удовлетворительную калибрационную способность. Однако аппроксимирующая диагональ, построенная с помощью MPM II 0, в большей степени соответствует графику наблюдаемой летальности. Модели APACHE II, APACHE IV и POSSUM обладают неудовлетворительной калибрационной способностью, что ограничивает возможности их использования для прогнозирования летальности в условиях ОРИТ хирургического профиля.

Прогностические шкалы APACHE II, APACHE IV, SAPS II, MPM II 0 имеют хорошую дискриминационную способность (площадь под ROC-кривой > 0,9), а шкала POSSUM удовлетворительную дискриминационную способность (площадь под ROC-кривой > 0,8). Калибрационная и дискриминационная способности прогностической системы являются комплементарными, то есть взаимно





дополняют друг друга и составляют единую систему характеристик, в которой неправдоподобность одного параметра означает неправдоподобность всей системы [10].

Таким образом, по данным нашего исследования, только две из пяти прогностических шкал (SAPS II и MPM II 0) могут быть рекомендованы для

оценки вероятности летального исхода в отделении реанимации хирургического профиля. Шкалы SAPS II и MPM II 0 обладают достаточной простотой, используют относительно небольшой набор доступных количественных и качественных параметров и могут успешно использоваться в повседневной практике отечественных ОРИТ.

ЛИТЕРАТУРА



1. Гельфанд Б.Р., Ярошецкий А.И., Проценко Д.Н., Романовский Ю.Я. Интегральные системы оценки тяжести состояния больных при политравме // Вестник интенсивной терапии. — 2004. — №1. — С. 1–10.
2. Евтюков Г.М., Александрович Ю.С., Иванов Д.О. Оценка тяжести состояния больных, находящихся в критическом состоянии // «Перспективы и пути развития неотложной педиатрии». Посвящается 120-летию первому в России обществу детских врачей. Санкт-Петербург, 3–4 февраля 2006 г.
3. Габдулхаков Р.М. Прогнозирование исходов и интенсивная терапия при сочетанной травме // Автореф. дис. ... д-ра мед. наук : 14.00.37, 14.00.27 / Башкирский государственный медицинский университет Федерального агентства по здравоохранению и социальному развитию. — М., 2009. — 47 с.
4. Лебедев Н.В., Корольков А.Ю. Системы объективных оценок тяжести состояния больных панкреонекрозом // Хирургия. Журнал им. Н.И. Пирогова. — 2005. — №7.
5. Сотников А.В., Кормуков И.А. Прогностическое значение шкал SAPS II и APACHE II у больных с синдромом полиорганной недостаточности раннего послеоперационного периода после онкохирургических вмешательств // Анестезиология и реаниматология. — 2003. — № 2. — С. 37–39.
6. Лutfарахманов И.И., Тимербuлатов В.М., Ишмухаметов И.Х., Миронов П.И. Прогностическая значимость систем оценки тяжести состояния при остром панкреатите // Анналы хирургии: Научно-практический журнал. — 2005. — № 5. — С. 38–42.
7. Зарубина Т.В., Швырев С.Л., Сидоров К.В. ИНТЕРИС — информационная система отделения реанимации и интенсивной терапии // Врач и информационные технологии. — 2006. — № 3. — С. 27–40.
8. Knaus W.A., Wagner D.P., Draper E.A., Zimmerman J.E., Bergner M., Bastos P.G., Sirio C.A., Murphy D.J., Lotring T., Damiano A. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults // Chest. — 1991. — Vol. 100. — P. 1619–1636.
9. Le Gall J.R., Lemeshow S., Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study // JAMA. — 1993. — Vol. 270. — P. 2957–2963.
10. Lemeshow S., Teres D., Klar J., Avrunin J.S., Gehlbach S.H., Rapoport J. Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients // JAMA. — 1993. — Vol. 270. — P. 2478–2486.
11. Fagon J.Y., Chastre J., Novara A., Medioni P., Gibert C. Characterization of intensive care unit patients using a model based on the presence or absence of organ dysfunctions and/or infection: the ODIN model // Intensive Care Med. — 1993. — № 19. — P. 137–144.



12. *Nashef SA, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R.* European system for cardiac operative risk evaluation (EuroSCORE)//*Eur. J. Cardiothorac. Surg.* — 1999. — № 16. — P. 9–13.
13. *Tu J.V., Jaglal S.B., Naylor C.D.* Multicenter validation of a risk index for mortality, intensive care unit stay, and overall hospital length of stay after cardiac surgery//*Circulation.* — 1995. — № 91. — P. 677–684.
14. *Groeger J.S., Lemeshow S., Price K., Nierman D.M., White P., Klar J., Granovsky S., Horak D., Kish S.K.* Multicenter outcome study of cancer patients admitted to the Intensive Care Unit: a probability of mortality Model//*J Clin Oncol.* — 1998. — № 16. — P. 761–770.
15. *Metz C.E.* Basic principles of ROC analysis//*Seminars in Nuclear Medicine.* — 1978. — № 8. — P. 283–298.
16. *Zweig M.H., Campbell G.* Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine//*Clinical Chemistry.* — 1993. — № 39. — P. 561–577.
17. The GNU project: a language and environment for statistical computing and graphics. URL: <http://www.r-project.org/> (Дата обращения 24.12.2010).
18. *Griner P.F., Mayewski R.J., Mushlin A.I., Greenland P.* Selection and interpretation of diagnostic tests and procedures//*Annals of Internal Medicine.* — 1981. — № 94. — P. 555–600.
19. *Mourouga P., Goldfrad C., Rowan K.M.* Does it fit? Is it good? Assessment of scoring systems//*Current Opinion in Critical Care.* — 2000. — № 6. — P. 176–180.
20. *DeLong E.R., DeLong D.M., Clarke-Pearson D.L.* Comparing the areas under two or more correlated receiver operation characteristic curves: a nonparametric approach//*Biometrics.* — 1988. — Sep. — № 44(3). — P. 837–945.

Медицинские ИТ-инновации



ИНТЕЛЛЕКТУАЛЬНОЕ БЕЛЬЕ

Профессор кафедры патологии Университета штата Вирджиния Робин Фелдер (Robin Felder) прогнозирует: «В ближайшем будущем в нижнем белье и других предметах повседневной одежды появятся встроенные датчики кровяного давления и пульса... По крайне невысокой себестоимости фармацевтические компании смогут добавлять в таблетки безопасные для пищеварительного тракта микросхемы, которые будут фиксировать факт приема препарата пациентом, измерять кислотность желудочного сока и другие важнейшие показатели и передавать данные на сотовый телефон через Bluetooth».

Фелдер работал с компаниями-разработчиками контактных линз и других глазных аппликаторов с датчиками, автоматически измеряющими уровень глюкозы в выделениях слезных желез диабетических пациентов и передающими полученные данные на сотовый телефон. Однако путевку этим технологиям в жизнь дают не столько новые устройства, сколько активное включение телекоммуникационных компаний в экосистему телемедицины.

Источник: Mobile Health News