

Прогностическая эффективность биомаркеров

Рубанович А.В.

*Институт общей генетики им. Н.И. Вавилова РАН,
лаборатория экологической генетики*

rubanovich@vigg.ru

Прогностическая эффективность биомаркеров

ИЛИ

как представить результаты так,
чтобы они нравились не только нам,
но и в министерствах

ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007—2012 годы»

Лот № 2. 2011-1.2-512-014. Разработка постгеномных методов для молекулярно-генетической диагностики соматических заболеваний

17	НИИ общей патологии и патофизиологии РАМН	<p>Разрабатываемая молекулярно-диагностическая тест-система для ранней диагностики генетической предрасположенности к развитию особо опасных и распространенных заболеваний обеспечит следующие диагностические характеристики:</p> <ol style="list-style-type: none"> 1. специфичность - не менее 80%; 2. чувствительность - не менее 83%; 3. точность - не менее 90%;
----	---	--

22	ГНИИ генетики и селекции промышленных микроорганизмов	<p>высокий (до 98%) уровень информативности (чувствительность/специфичность) создаваемых систем полиморфных маркеров, основанных на идентификации этиологических вариантов генов-кандидатов нового способа диагностики му диабету типа 1 (95% или более).</p>
----	---	---

О чем это они?

32	Российский онкологический научный центр имени Н.Н. Блохина РАМН	<p>значимых полиморфных вариантов, вовлеченных в канцерогенез наследственных и спорадических форм рака молочной железы, рака яичников и рака щитовидной железы;</p> <p>-определение чувствительности и специфичности применяемых методик</p>
----	---	--

Лот № 6. 2011-1.2-512-027. Исследование вариантов генов, ассоциированных с устойчивостью и чувствительностью человека к воздействию радиации

57	ГНИИ генетики и селекции промышленных микроорганизмов	<p>Высокий (до 98%) уровень информативности (чувствительность/специфичность) создаваемых молекулярно-генетических систем, основанных на идентификации вариантов генов-кандидатов. Высокая чувствительность нового способа диагностики устойчивости человека к воздействию радиации (95% или более).</p>
----	---	---

Чувствительность – специфичность: старые добрые медицинские понятия

Чувствительность (SE) – это доля носителей маркера среди больных (тест сработал!)

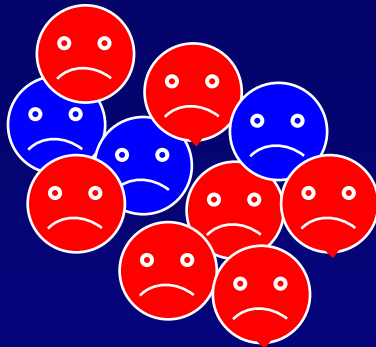
$$SE = P(M|D)$$

Специфичность (SP) – это доля здоровых, которые не являются носителем маркера (тест сработал!)

$$SP = P(\bar{M}|\bar{D})$$

Случаи, когда
маркер сработал

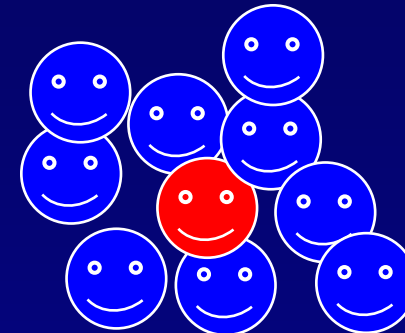
10 больных



$$OR = \frac{7.9}{3.1} = 21$$

при $p = 0.02$

10 здоровых



Чувствит

Случаи, когда
маркер не сработал

Специфичность = 0.9

SE vs. *SP*: противоборство показателей

Все без маркера,
и все здоровы!
 $SE=0, SP=1$



Маркер у всех,
и все больны!
 $SE=1, SP=0$

Выигрывая в чувствительности,
обычно теряем специфичность (*et converso*)

Так что важнее: чувствительность или специфичность?



Тест должен быть высокочувствительным, если важно не пропустить ни одного больного (пусть даже будет гипердиагностика).



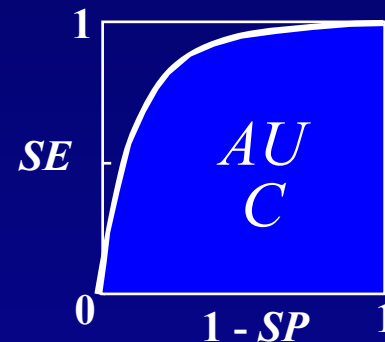
Тест должен быть высокоспецифичным, если важно не объявлять здорового больным (серьезная психологическая травма либо сильные побочные эффекты лечения – гипердиагностика нежелательна)



Все зависит от ситуации, но кроме того существует простой и эффективный показатель прогностических возможностей маркера (отличать больного от здорового):

$$AUC = \frac{SE + SP}{2}$$

Area Under Curve?



Почему величина $AUC = (SE + SP) / 2$ характеризует прогностическую эффективность маркера?

Берем
1 здорового и 1 больного

□ Без использования маркера



вероятность угадать,
кто есть кто = $1/2$

□ Используем тест и носителя маркера объявляем больным



вероятность угадать,
кто есть кто = $AUC = \frac{1}{2}SE + \frac{1}{2}SP$

В качестве больного выбираем:



- любого

AUC — это вероятность отличить **любого** больного от здорового, ориентируясь на маркер **- с маркером**

$AUC > 1/2$, если $OR > 1$

AUC	Прогностическая ценность
0.5	Случайный классификатор
0.5-0.6	Плохой классификатор
0.6-0.7	Средний классификатор
0.7-0.8	Хороший классификатор
>0.8	Отличный классификатор

А зачем все это?

REVIEW ARTICLE

Относительная полезность OR и RR

Ограниченная полезность OR при использовании диагностических и прогностических маркеров

Маркеры с воспроизводимыми значимыми OR могут быть плохими классификаторами

Poor Classifier

Johanna Jakobsdottir

Invited critical review

...брешь между статистической значимостью и диагностической дискриминацией

diagnostic discrimination

Abstract | The brisk discovery of novel inherited disease markers by genome-wide association (GWA) studies has raised expectations for predicting disease risk by

По ту сторону OR – риск заболеваемости на основе генетических профилей

'clinical validity', namely the sensitivity, specificity, and positive and negative predictive values associated with a genetic profile¹³. These parameters, defined in more detail below, summarize the ability of tests that are based on the genetic profile to discriminate between those who will develop disease and those who will not, as well as their ability to predict individual risk¹⁴.

We focus on clinical validity and underscore two points. First, measures of marker-disease association typically reported in the discovery phase of genet-

ness clinical WA studies measures of se data are ing to design rse' (the ation that ict have to

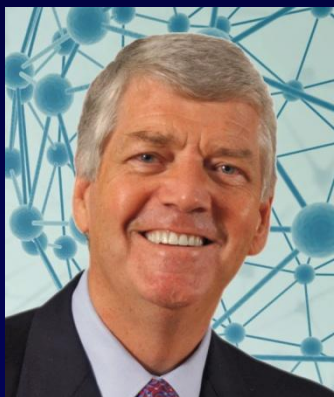
Прогноз генетического риска – умеем ли мы уже это делать ?

Peter Kraft, Ph.D., and David J. Hunter, M.B., B.S., Sc.D., M.P.H.

A major goal of the Human Genome Project was to facilitate the identification of inherit-

tests of genetic predisposition to important diseases would have major clinical, social, and econom-

est relative risks are almost certainly overrepresented in the first wave of findings from genome-



George Poste, *chief scientist at the Complex Adaptive Systems Initiative, and Regents professor of health innovation at Arizona State University*

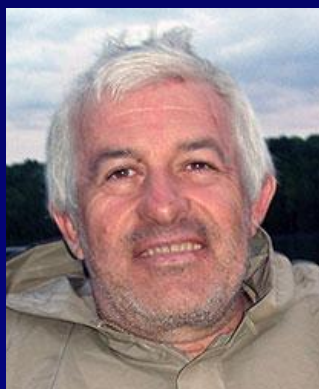
Нашествие биомаркеров

Nature , v 469, 2011

Не более 100

150 000

Good association is usually a necessary, but not a sufficient, condition for good classification



Проф. Алексей Кондрашов,
Life Sciences Institute, University of Michigan
Мегагрант-2010 (МГУ). Лекция в Полит.РУ (2010)
«Эволюционная биология человека и охрана здоровья»:

Надежды 90-х не оправдались: эпоха GWAS буксует

«У меня 10-15 генов предрасположенности?
В этой информации очень мало проку!»

Причина парадокса в низкой чувствительности теста:
 $50/500=0.1$, т.е. только 10% больных
 являются носителями маркера

Популяционное обследование:

1000 человек

Заболеваемость – 50%

500 больных

500 здоровых

Суперэффект!

50 с маркером

450 без маркера

5 с маркером

495 без маркера

$$AUC = \frac{1}{2} (SE + SP) = \frac{1}{2} \left(\frac{50}{500} + \frac{495}{500} \right) = \frac{1}{2} (0.1 + 0.99) = 0.545$$

Сколько раз тест сработал? Вероятность того, что случайно
 выбранный человек будет правильно идентифицирован равна:



$$\frac{50 + 495}{1000} = 0.545$$

- почти монета!!!
 Как и было



Разберемся в исходных понятиях:

все начинается с таблицы сопряженности 2×2:

		Больные Case	Здоровые Control
		D	\bar{D}
Носители маркера	M	a	b
Свободны от маркера	\bar{M}	c	d

Популяционная частота маркера

$$\text{Чувствительность} = SE = P(M | D) = \frac{a}{a+c} > p_M$$

$$\text{Специфичность} = SP = P(\bar{M} | \bar{D}) = \frac{d}{b+d} > 1-p_M$$

Интерпретация в терминах ошибок I и II рода

		Больные Case	Здоровые Control
	Тест сработал!	D	D
Носители маркера	M	True Positives	False Positive
Свободны от маркера	M	False Negative	True Negative

Тест не сработал: ошибка I рода

Тест сработал!

Тест не сработал: ошибка II рода

Плоховато запоминается - лучше частота маркера среди больных



$$SE = \frac{TP}{TP + FN}$$

$$SP = \frac{TN}{FP + TN}$$

Бинарный тест: вычисление показателей ассоциирования и риска

- Показатели ассоциирования
- Показатели рисков:

Доля дур среди блондинок, это не то же самое, что доля блондинок среди дур!

Во всех случаях можно оценить только относительные шансы



	D	\bar{D}	
M	a	b	$P(D M) = \frac{a+b}{a+b+c+d}$
\bar{M}	c	d	

«Абсолютные» риски:

доля больных среди M и \bar{M}

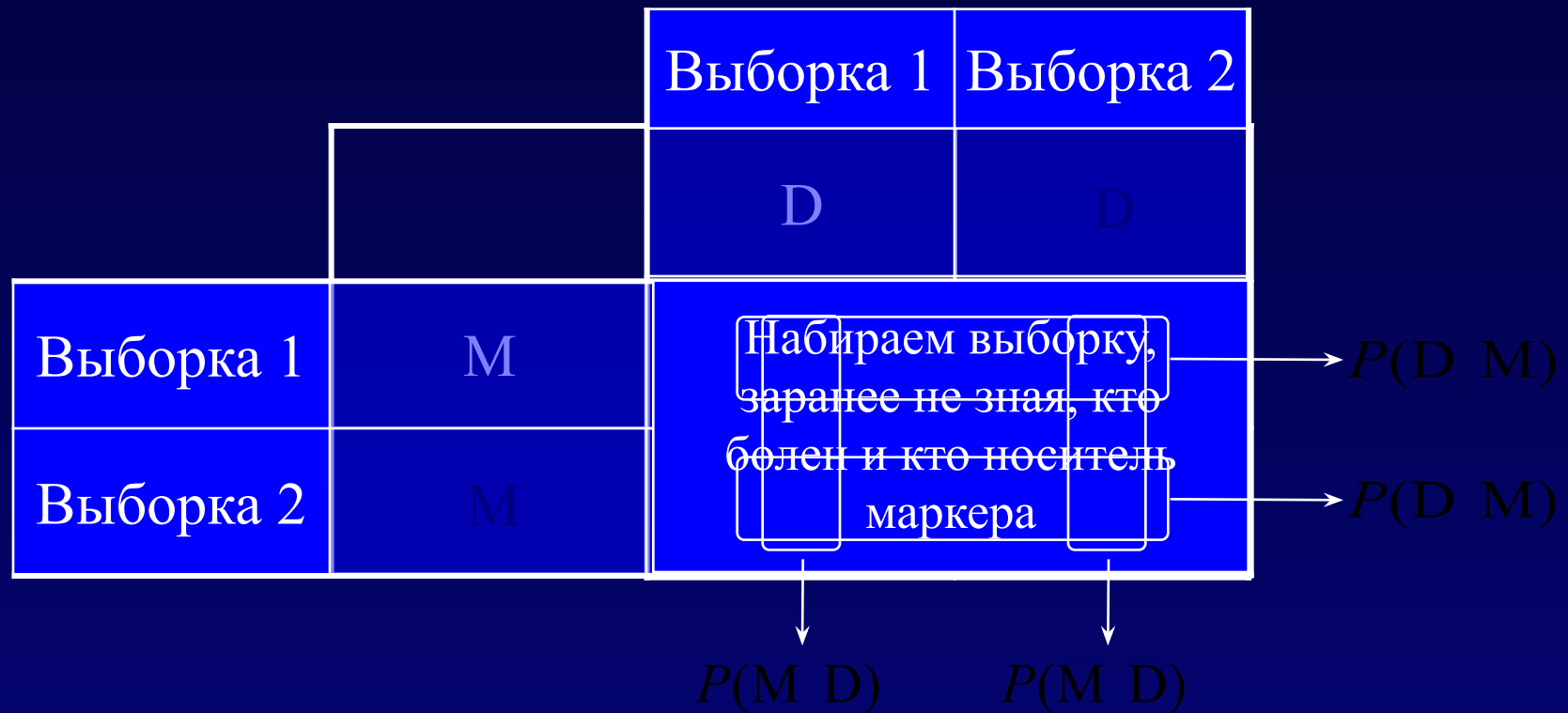
$$SE = P(M|D) = \frac{a}{a+c} \quad SP = P(\bar{M}|\bar{D}) = \frac{d}{b+d}$$

Чувствительность и специфичность:

Относительный риск:

Не все эти показатели поддаются непосредственной оценке – все зависит от дизайна эксперимента

Дизайн ассоциативных исследований



Case-control study:
подбор групп «больные-здоровые»

Основной вариант ассоциативных исследований



Cohort study:
подбор групп «с маркером – без маркера»

Обычно «экспонированные - контроль»



Population study:
случайная выборка без подбора групп

Редкий вариант (дорого!)

Возможности непосредственной оценки зависят

p_D – встречаемость (распространенность) заболевания

p_M – популяционная встречаемость (частота) маркера

Study	OR	RR	SE	SP	p_D	p_M
Case-control	✓	?	✓	✓	?	?
Cohort	✓	✓				
Popula						

При case-control мы не можем оценить заболеваемость и частоту маркера непосредственно в своем эксперименте

Однако в большинстве случаев в отношении p_D и p_M имеется априорная информация (эпидемиологические справочники, базы данных PubMed, NapMap и т.д.)

Зная OR , p_D и p_M можно оценить все остальные показатели (например, относительный риск RR для case-control)

На что можно рассчитывать, используя бинарный маркера?

Какими могут быть
для бинарного

При $OR > 1$:
 $SE > p_M$ и $SP > 1 - p_M$

специфичность и риски
в крайности:

😊 Идеальный маркер

	D	\bar{D}
M	a	b
\bar{M}	c	d

Чувствительность = $SE = P(M | D) = 1$

Специфичность = $SP = P(\bar{M} | \bar{D}) = 1$

$OR = RR = \infty$

$AUC = 1$

☹ Маркер не

В каждом
произведении
«встретить –

Чувствительность и
специфичность больше нуля даже
в отсутствии ассоциаций

	D	\bar{D}
M	$p_M p_D$	$p_M (1 - p_D)$
\bar{M}	$(1 - p_M) p_D$	$(1 - p_M) (1 - p_D)$

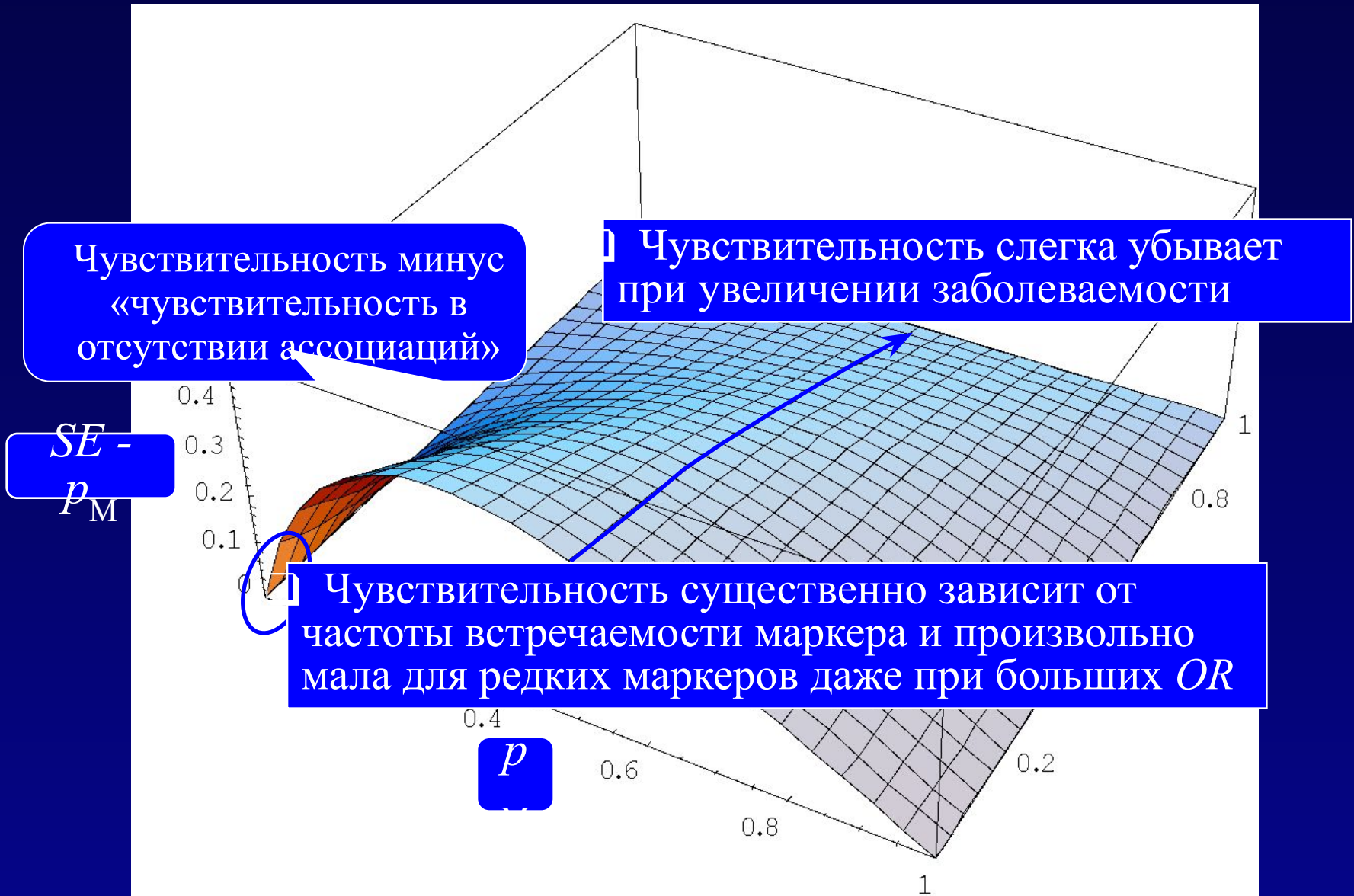
Чувствительность = $SE = P(M | D) = p_M$

Специфичность = $SP = P(\bar{M} | \bar{D}) = 1 - p_M$

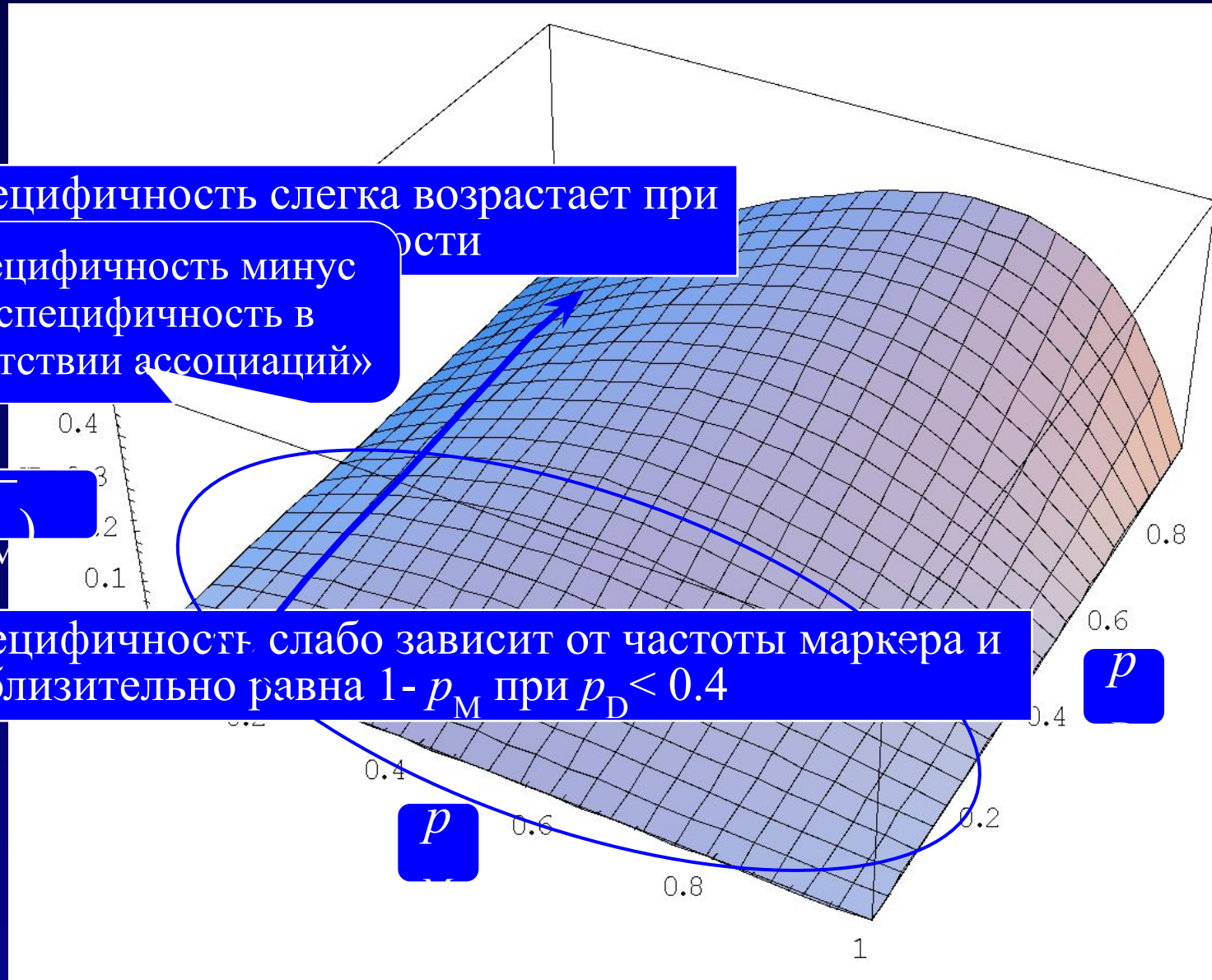
$OR = RR = 1$

$AUC = 1/2$

Зависимость чувствительности от p_D и p_M при фиксированном $OR (=5)$



Зависимость специфичности от p_D и p_M при фиксированном $OR(=5)$



□ Специфичность слегка возрастает при

росту

Специфичность минус
«специфичность в
отсутствии ассоциаций»

$$SP = 1 - p_M$$

□ Специфичность слабо зависит от частоты маркера и приблизительно равна $1 - p_M$ при $p_D < 0.4$

p

p

Зависимость AUC от p_D и p_M

Величина максимально возможного AUC не зависит от p_M и p_D и равна $\frac{\sqrt{OR}}{\sqrt{OR}+1}$

$$\frac{OR}{OR+1}$$

AUC

$OR=20$

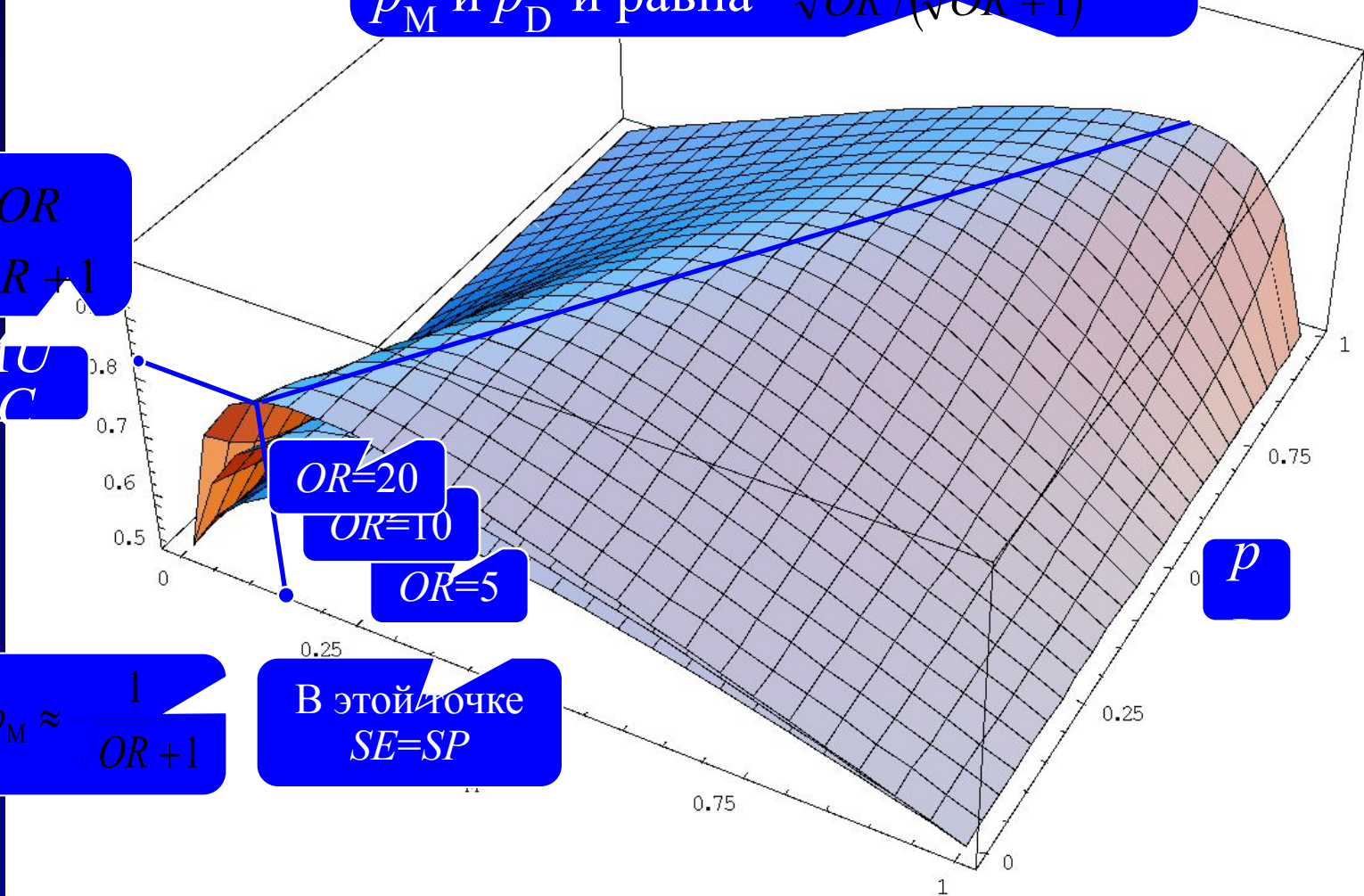
$OR=10$

$OR=5$

p

$$p_M \approx \frac{1}{OR+1}$$

В этой точке $SE=SP$



При каких OR маркер является хорошим классификатором?

☺ Когда $OR > 5$, а частота маркера не менее 30%:

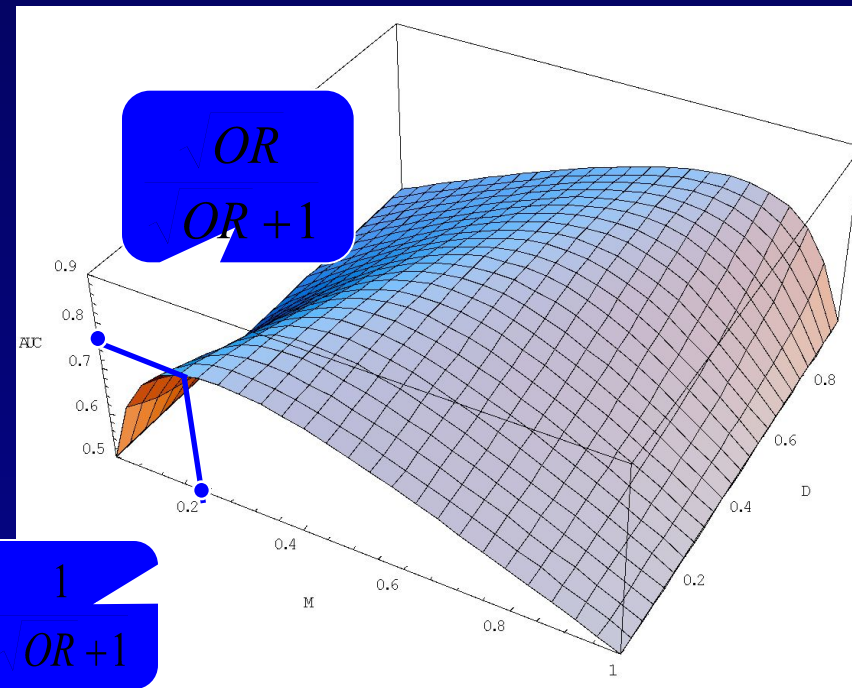
$$\max AUC = \frac{OR}{OR + 1} > 0.7 \implies OR > 5.4 \text{ при } p_M \approx \frac{1}{\sqrt{OR + 1}} = \frac{1}{\sqrt{5.44 + 1}} \approx 0.3$$

☹ При низких $OR (< 2.5)$ маркер является заведомо плохим классификатором при любых частотах заболевания и маркера:

$$\max AUC = \frac{OR}{OR + 1} < 0.6 \implies OR < 2.4$$

☹ При $p_M \cdot OR \ll 1$
чувствительность теста заведомо мала:

$$SE < p_M \cdot OR$$



Почему высокое OR не всегда гарантирует эффективность прогнозов, и когда подобное происходит?

При низкой частоте встречаемости маркера тест прогностически неэффективен. Точнее ...

даже при очень больших OR маркер является плохим классификатором, если $p_M / p_D < 0.2$ (тогда $AUC < 0.6$)

При $P(M | \bar{D}) = 0$ имеем: $OR = \infty$ и $p_D > p_M$

$$SE = P(M | D) = \frac{p_M}{p_D}$$

	D	D
M	50	5
M	450	495

$SP = P(M | D) = \frac{1}{1} = 1$

	D	D
M	p_M	0
M	$p_D - p_M$	$1 - p_D$

p_D $1 - p_D$

Мало

$$AUC = \frac{1}{2} \left(1 + \frac{p_M}{p_D} \right)$$

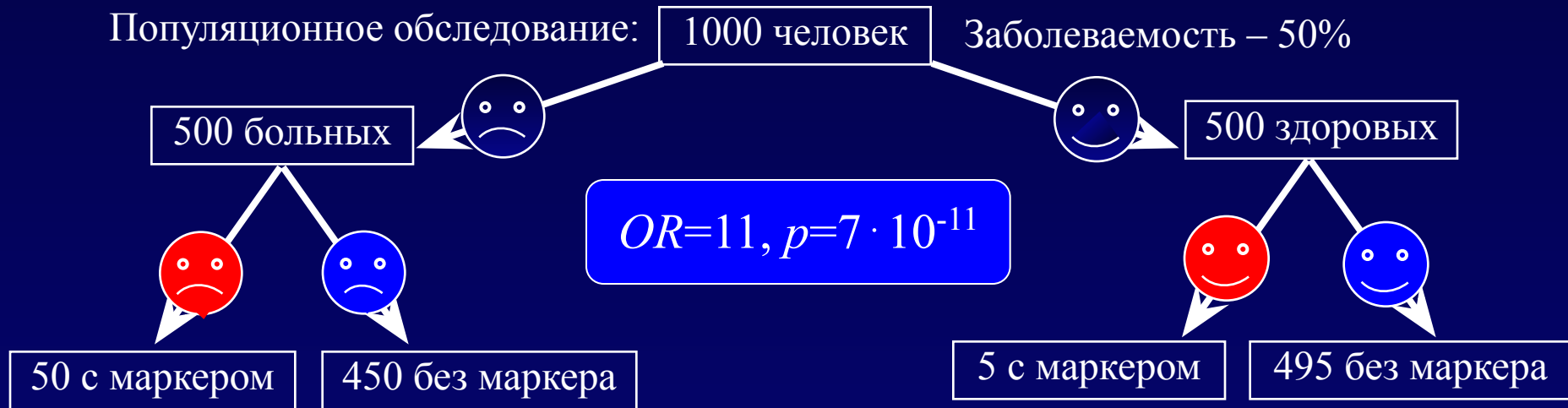
Возьмем крайний случай: $OR = \infty$

$AUC \rightarrow 1/2$
при $p_M \rightarrow 0$,
хотя $OR = \infty$

p_M
 $1 - p_M$

И еще одно обстоятельство

$OR=11$, но это не совсем относительный риск
(который показывает во сколько раз заболеваемость выше у носителей маркера)



$$\text{Доля больных среди носителей маркера} = P(D | M) = \frac{50}{50 + 5} = 0.91$$

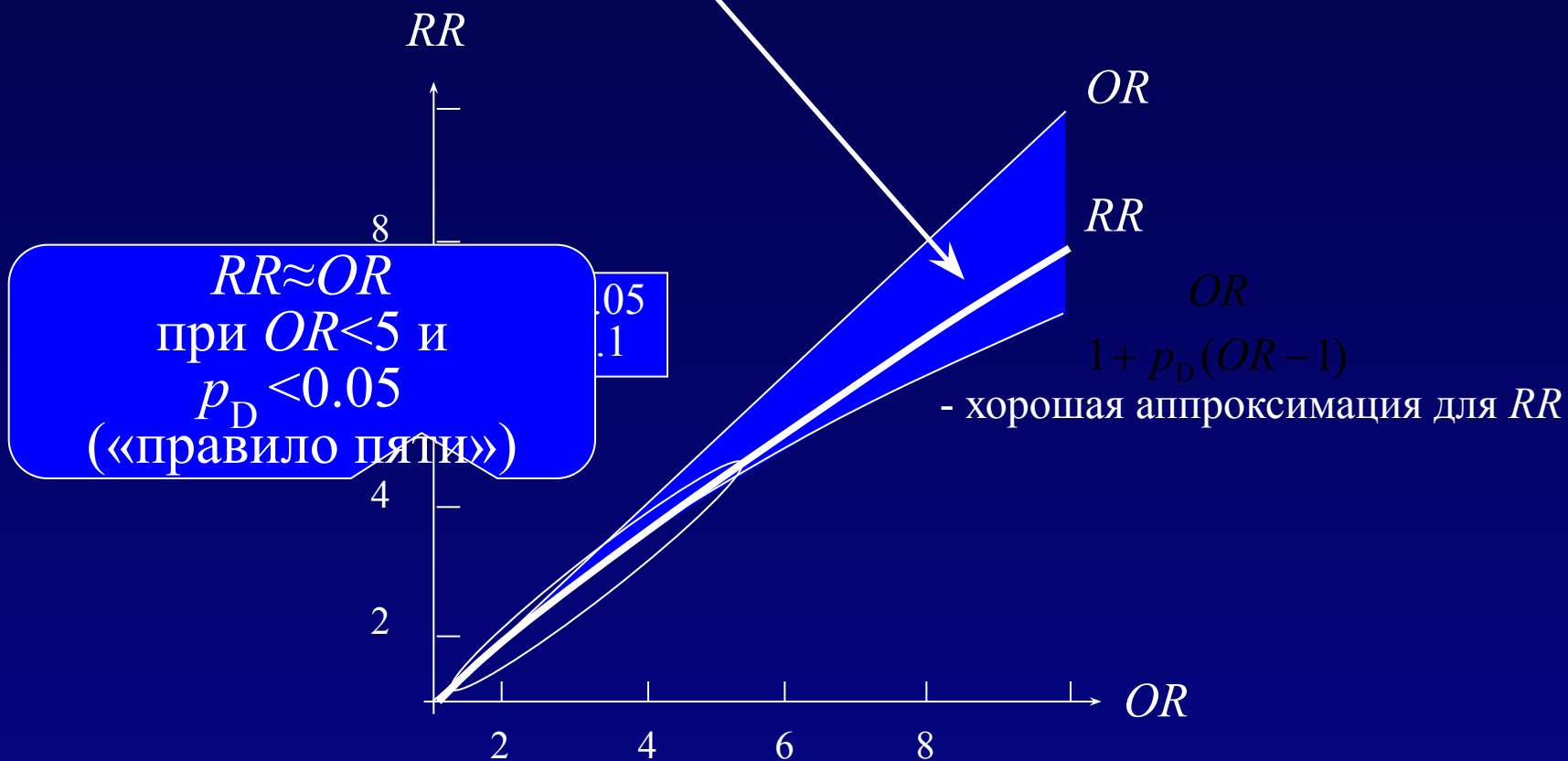
$$\text{Доля больных среди свободных от маркера} = P(D | \bar{M}) = \frac{450}{450 + 495} = 0.47$$

$$\text{Относительный риск} = RR = \frac{0.91}{0.47} \approx 1.9 \ll OR = 11$$

OR завышает оценку относительного риска

При $OR > 1$:

$$OR \geq \boxed{RR} \geq \frac{OR}{1 + p_D(OR - 1)} \quad \underline{\text{(всегда!)}}$$



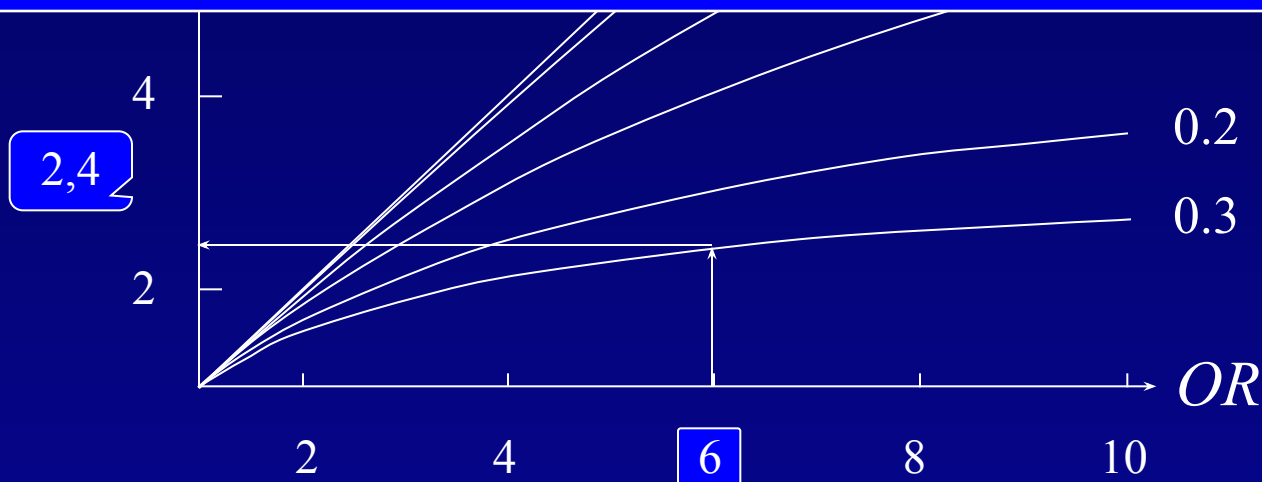
RR может быть существенно меньше OR

в случае широко распространенного заболевания ($p_D > 0.1$)
и редкого маркера ($p_M < 0.05$)



Если редкий ген (MAF=1%) сильно ассоциирован ($OR=6$) с широко распространенным заболеванием, то это не значит, что носители этого гена заболевают в 6 раз чаще

На самом деле только в 2.4 раза чаще



Бинарные и количественные маркеры (тесты)

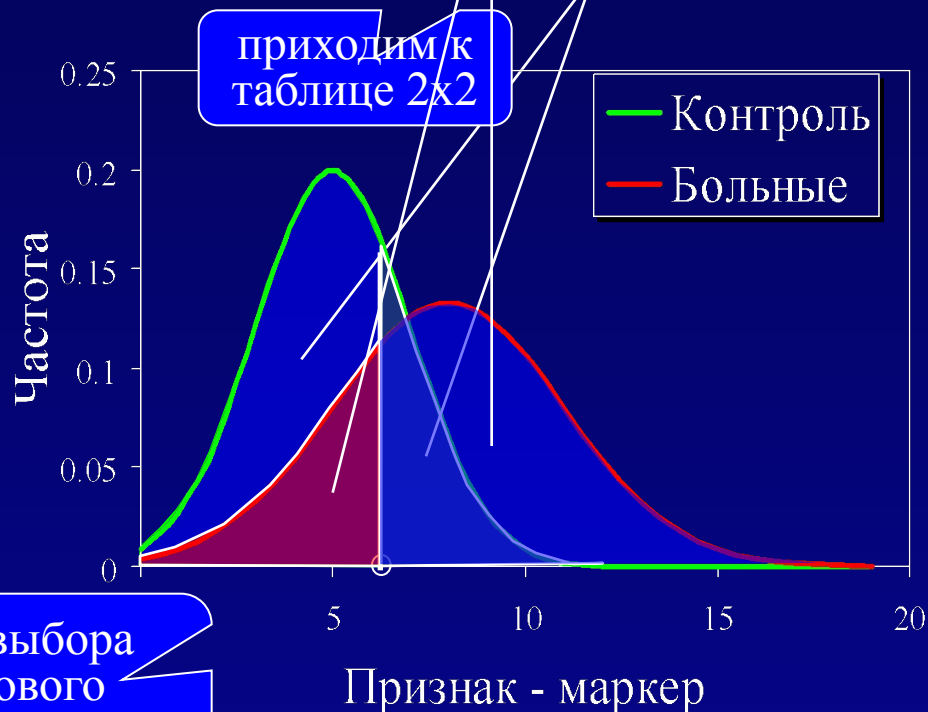
Бинарный тест:
маркер «есть-нет»



Количественный тест:
маркер является
количественным
показателем

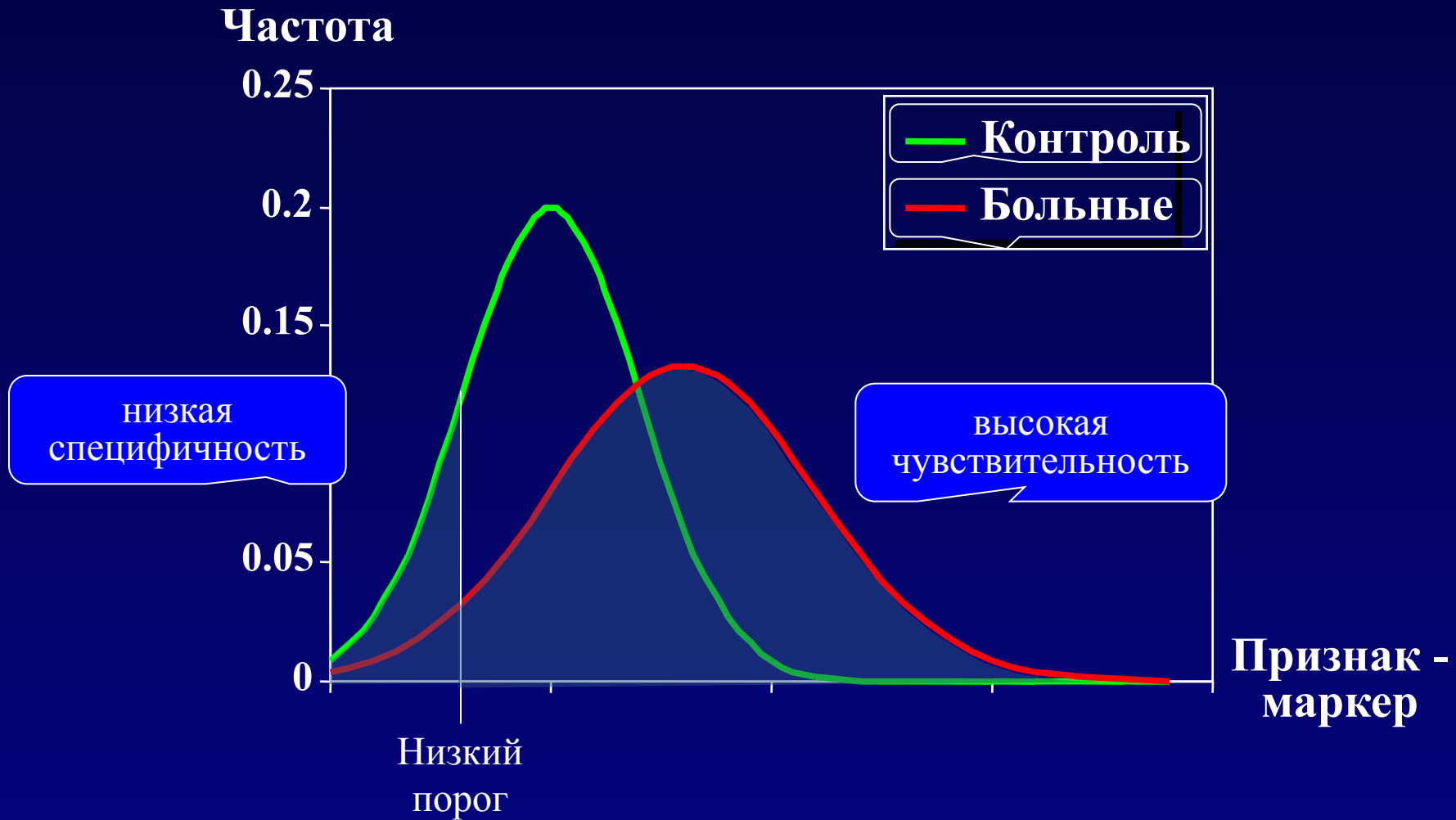


	D	D
M	a	b
M	c	d

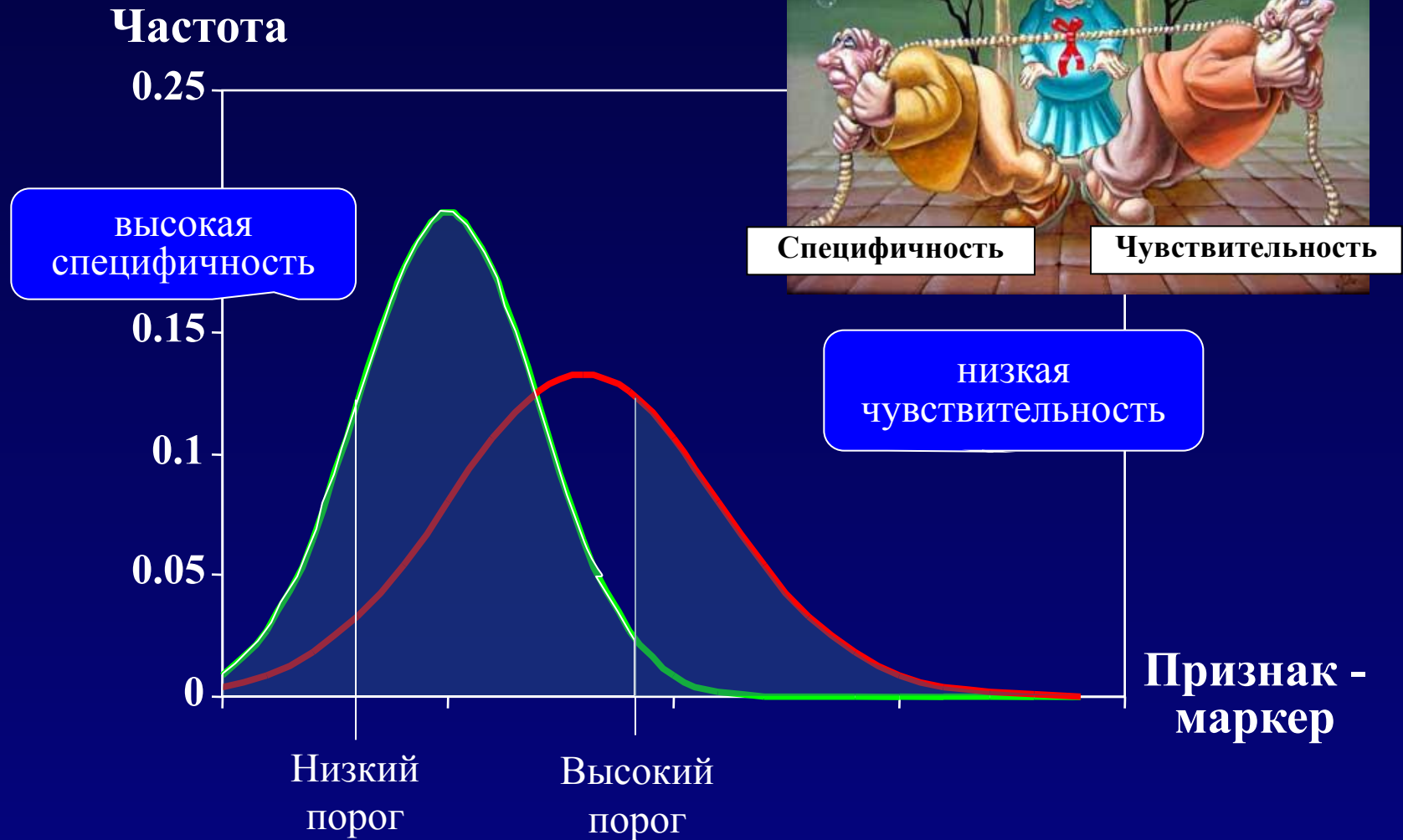


После выбора
порогового
значения маркера

Количественный тест: выбор порога



Количественный тест:



Как выбрать порог, чтобы соотношение чувствительность/специфичность было оптимальным?

ROC – анализ Receiver Operator Characteristics

Термин времен 2-ой мировой войны, который придумали операторы первых радарных установок

Как отличить шум и помехи от вражеского самолета?

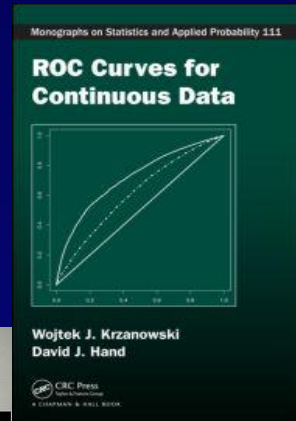
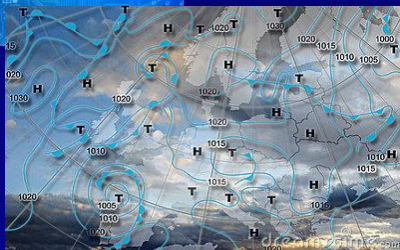
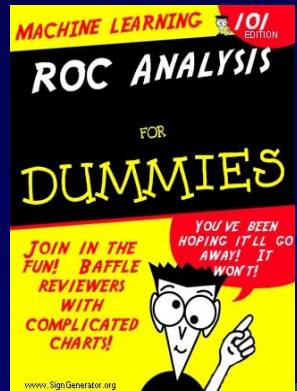
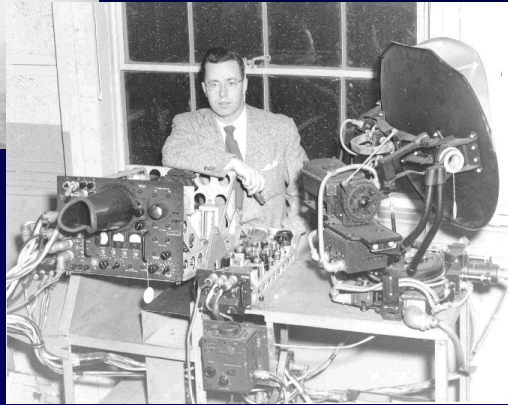
Современные применения ROC-анализа:

Медицинская диагностика, радиология и физиология

Распознавание образов, машинное обучение

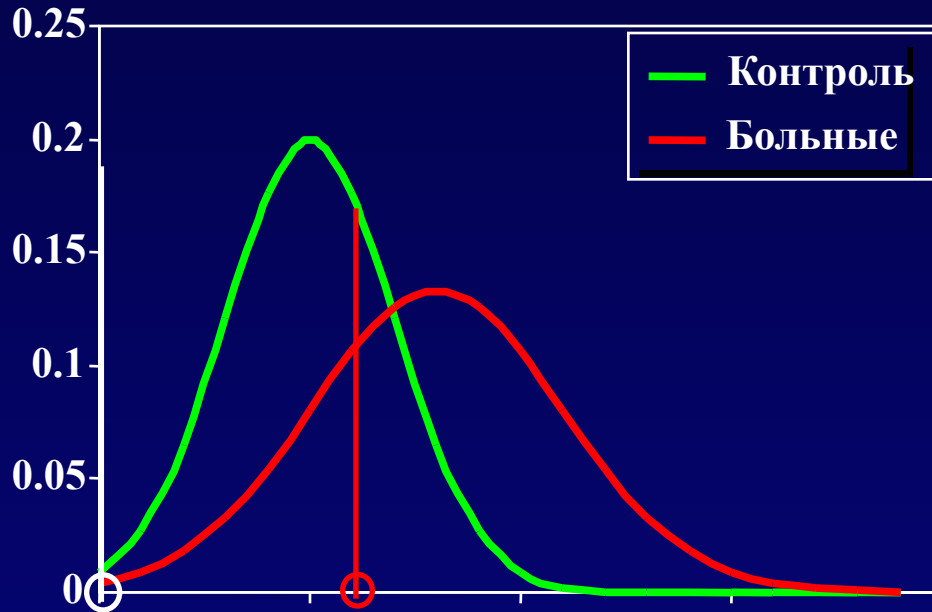
Погода и любые прогнозы по приметам

Борьба со спамом

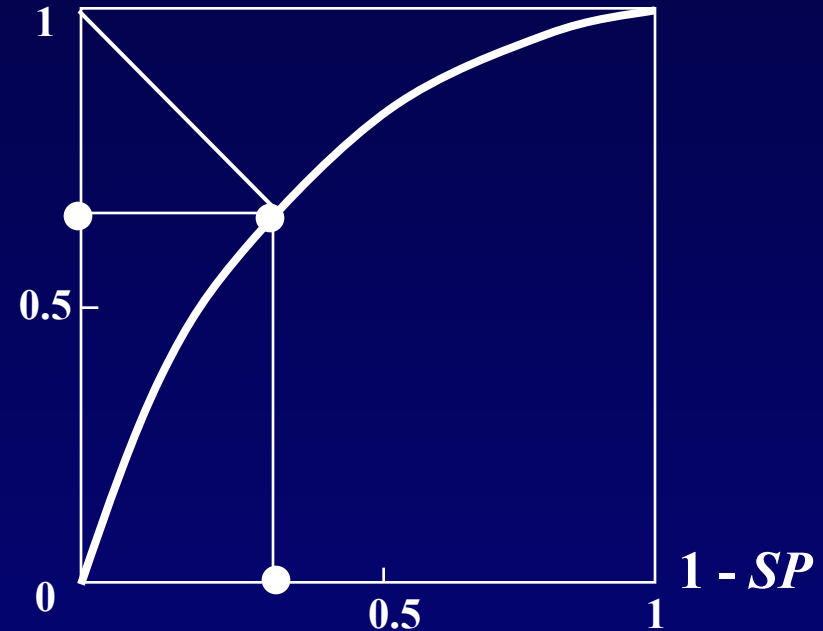


ROC – кривая: зависимость SE от $(1-SP)$ при плавном изменении порога

Частота



SE



Пороговое значение

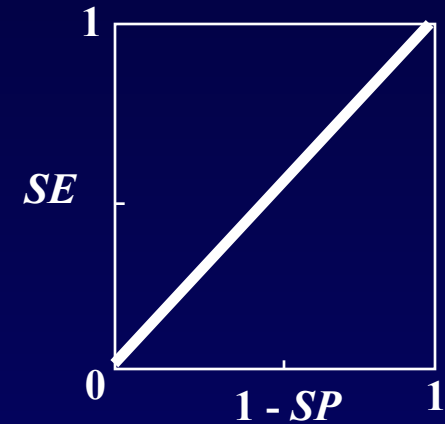
Признак - маркер

Оптимальный выбор порога: выбирается точка на ROC-кривой, которая ближе всех к левому верхнему углу $(0,1)$, т.е.

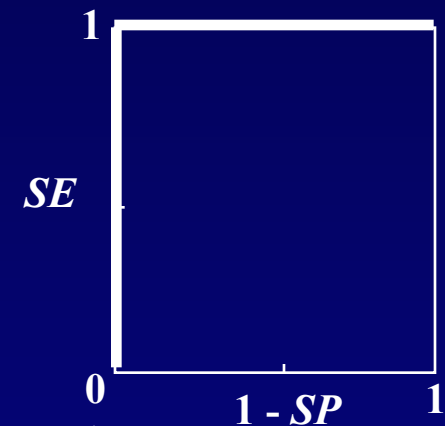
$$\sqrt{(1-SP)^2 + (1-SE)^2} \rightarrow \min$$

Форма ROC-кривых

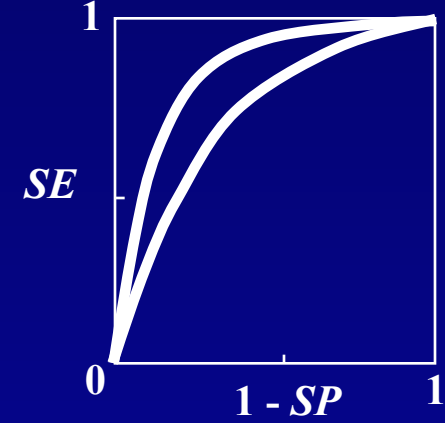
❑ Маркер не работает



❑ Идеальный маркер



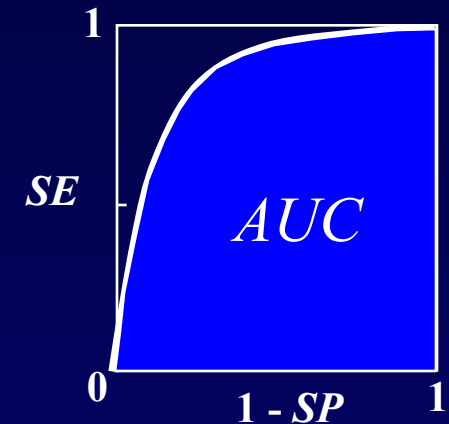
❑ 1-ый тест лучше, чем 2-ой



AUC – интегральный показатель прогностической эффективности маркера

AUC - это площадь под ROC-кривой
(**A**rea **U**nder **C**urve)

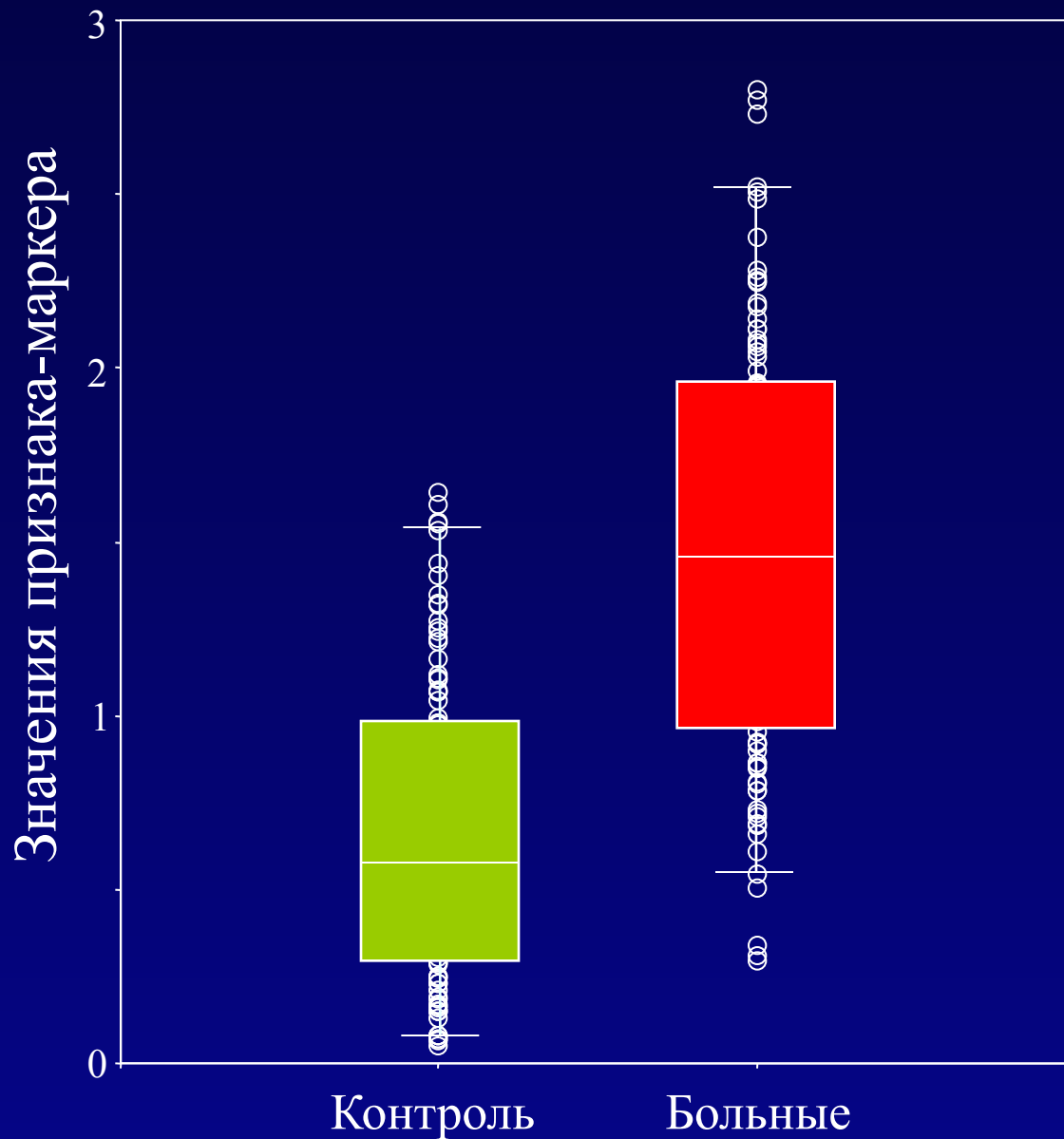
AUC = Вероятность того, что значение признака-маркера у случайно выбранного больного больше, чем у случайно выбранного здорового



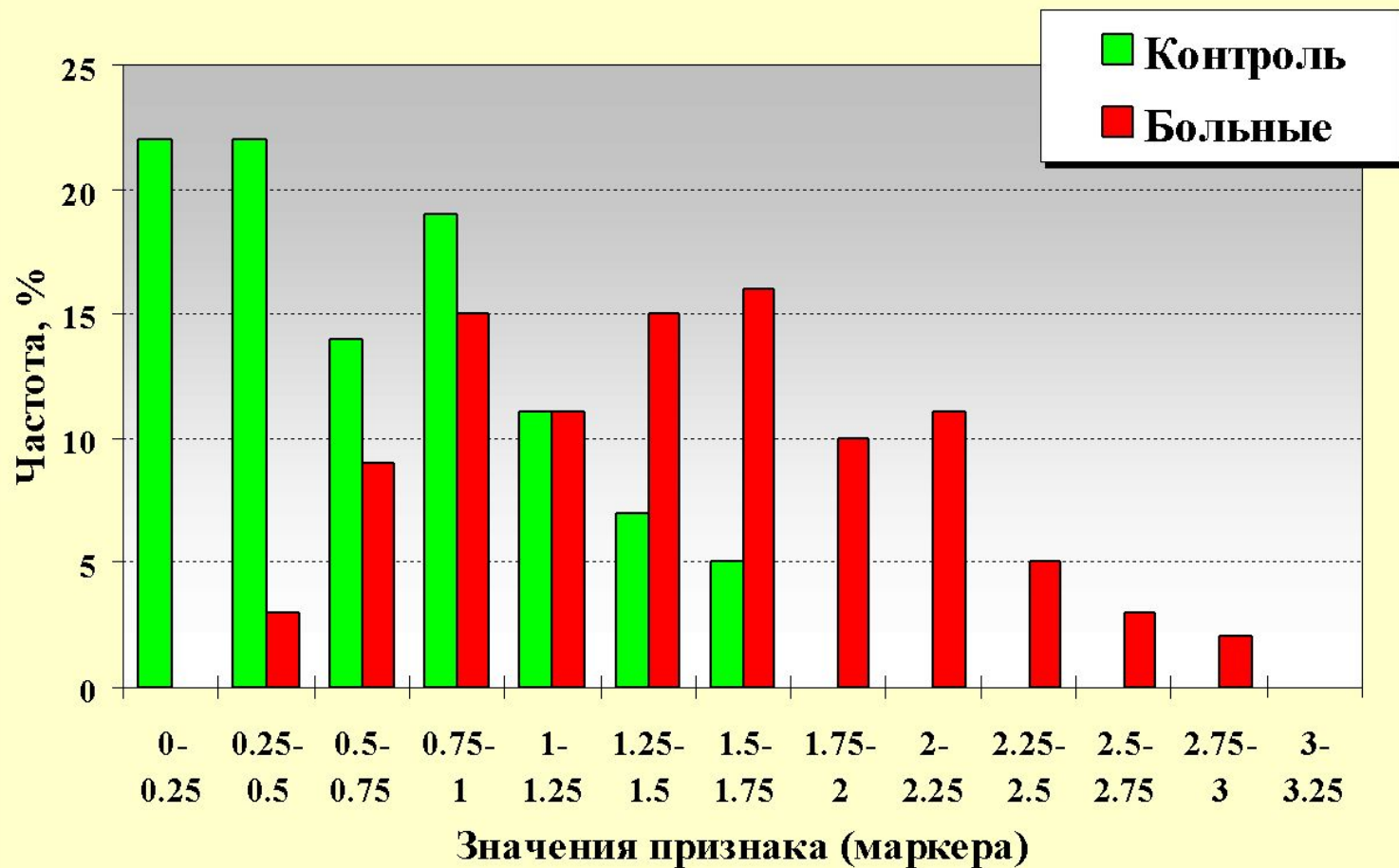
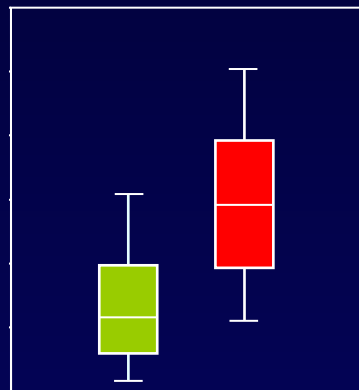
<i>AUC</i>	
0.5	Случайный классификатор
0.5-0.6	Плохой классификатор
0.6-0.7	Средний классификатор
0.7-0.8	Хороший классификатор
>0.8	Отличный классификатор

$U = N_1 \cdot N_2 \cdot AUC$ - непараметрический критерий Манна-Уитни

Вычисления при ROC-анализе



Вычисления при ROC-анализе

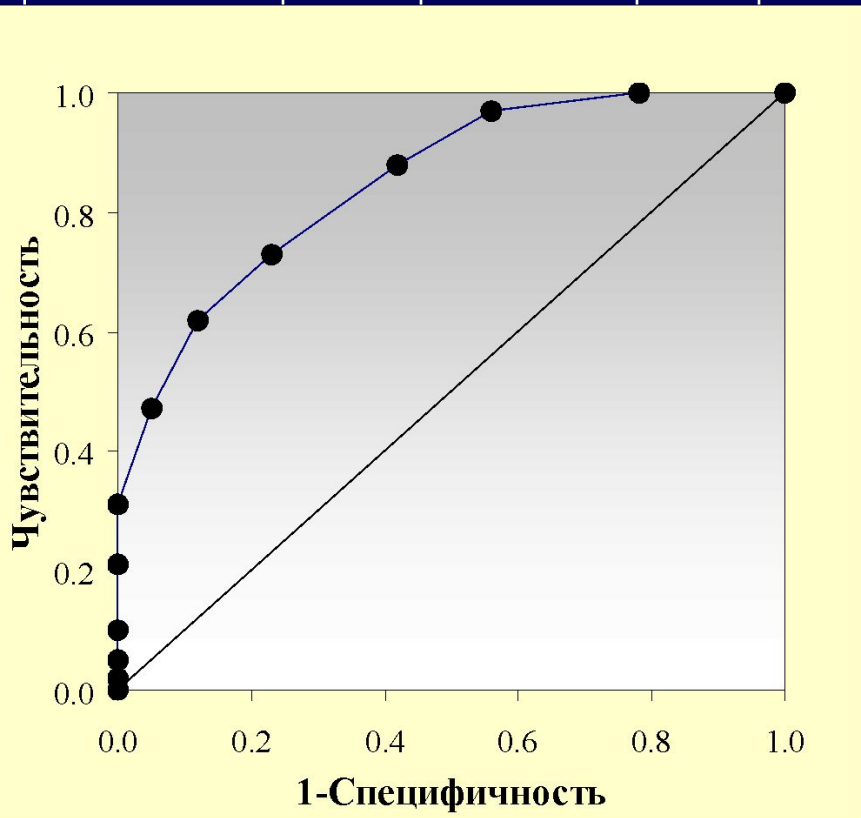


Все делается в Excel:

Интервал	Контроль		Больные		<i>SE</i>	<i>SP</i>
	#	Доля	#	Доля		
0-0.25	22	0.22	0	0		
0.25-0.5	22	0.22	3	0.03		
0.5-0.75	14	0.14	9	0.09	0.88	0.58
0.75-1	19	0.19	15	0.15		
1-1.25	11	0.11	11	0.11		
1.25-1.5	7	0.07	15	0.15		
1.5-1.75	5	0.05	16	0.16		
1.75-2	0	0	10	0.10		
2-2.25	0	0	11	0.11		
2.25-2.5	0	0	5	0.05		
2.5-2.75	0	0	3	0.03		
2.75-3	0	0	2	0.02		
3-3.25	0	0	0	0		

Все делается в Excel:

Интервал	Контроль		Больные		SE	SP
	#	Доля	#	Доля		
0-0.25	22	0.22	0	0	1	0,22
0.25-0.5	22	0.22	3	0.03	0,97	0,44
0.5-0.75	14	0.14	9	0.09	0,88	0,58

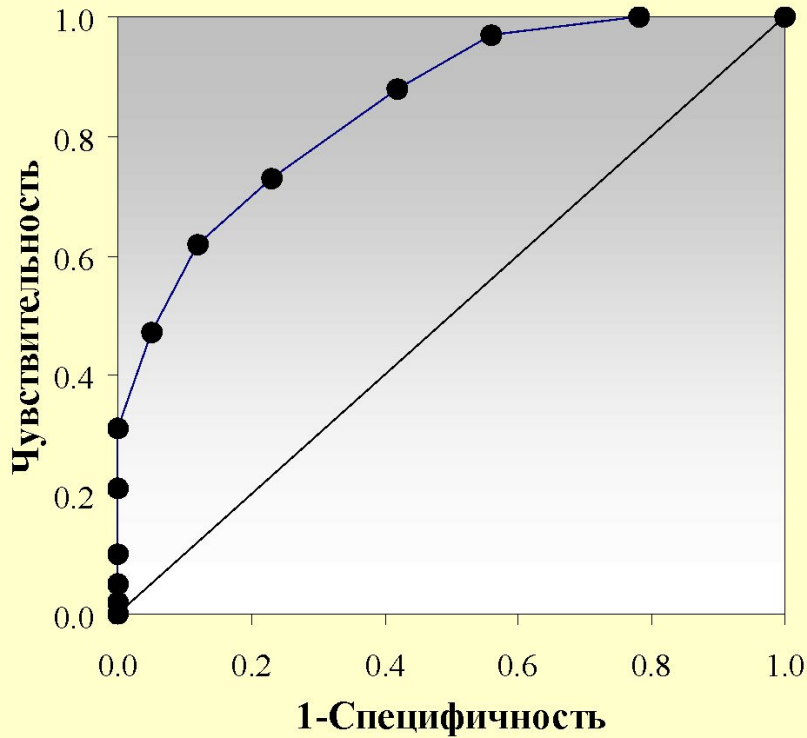


5	0,73	0,77
1	0,62	0,88
5	0,47	0,95
6	0,31	1
0	0,21	1
1	0,1	1
5	0,05	1
3	0,02	1
2	0	1
	0	1

Все делается в Excel:

Интервал	Контроль		Больные		SE	SP	$L_{(0,1)}$
	#	Доля	#	Доля			
0-0.25	22	0.22	0	0	1	0,22	
0.25-0.5	22	0.22	3	0.03	0.97	0.44	
0.5-0.75	14	0.14	3	0.03	0.97	0.58	

$$\sqrt{(1-SP)^2 + (1-SE)^2}$$

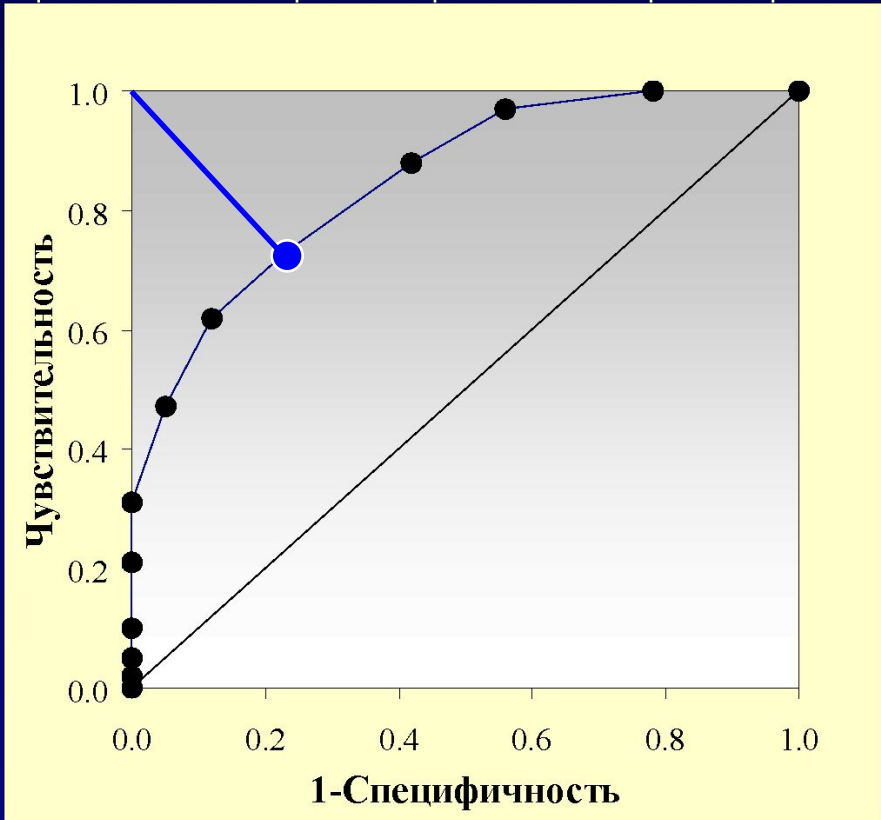


1	0,62	0,88
5	0,47	0,95
6	0,31	1
0	0,21	1
1	0,1	1
5	0,05	1
3	0,02	1
2	0	1
	0	1

Все делается в Excel:

Интервал	Контроль		Больные		SE	SP	$L_{(0,1)}$
	#	Доля	#	Доля			
0-0.25	22	0.22	0	0	1	0,22	0,780
0.25-0.5	22	0.22	3	0.03	0,97	0,44	0,561
0.5-0.75	14	0.14	9	0.09	0,88	0,58	0,437
					0,73	0,77	0,355
					0,62	0,88	0,398
					0,47	0,95	0,532
					0,31	1	0,690
					0,21	1	0,790
					0,1	1	0,900
					0,05	1	0,950
					0,02	1	0,980
					0	1	1,000
					0	1	1,000

Наименьшее расстояние до точки (0,1)



Все делается в Excel:

Интервал	Контроль		Больные		SE	SP	$L_{(0,1)}$
	#	Доля	#	Доля			
0-0.25			0	0	1	0,22	0,780
0.25-0.5			3	0.03	0,97	0,44	0,561
0.5-0.75	14	0.14	9	0.09	0,88	0,58	0,437
0.75-1	19	0.19	15	0.15	0,73	0,77	0,355
1-1.25	11	0.11	11	0.11	0,62	0,88	0,398
1.25-1.5					0,95		0,532
1.5-1.75					1		0,690
1.75-2					1		0,790
2-2.25					1		0,900
2.25-2.5	0	0	5	0.05	0,05	1	0,950
2.5-2.75	0	0	3	0.03	0,02	1	0,980
2.75-3	0	0	2	0.02	0	1	1,000
3-3.25	0	0	0	0	0	1	1,000

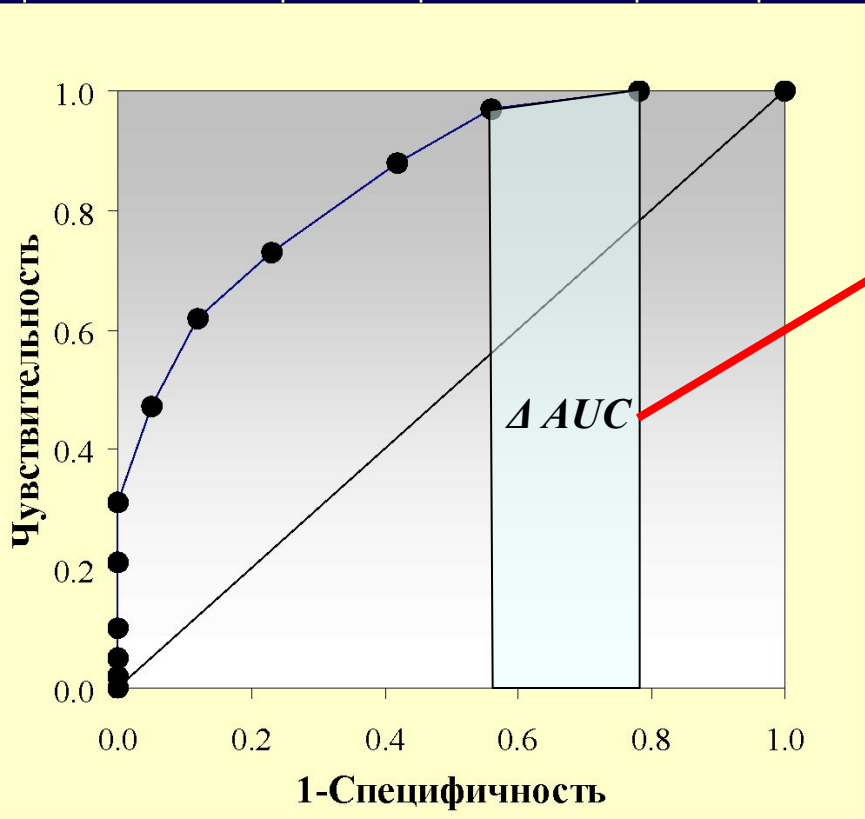
Оптимальное пороговое значение

Наименьшее расстояние до точки (0,1)

При выборе этого порога мы правильно идентифицируем 73% больных и 77% здоровых

Все делается в Excel:

Интервал	Контроль		Больные		SE	SP	$L_{(0,1)}$	ΔAUC
	#	Доля	#	Доля				
0-0.25	22	0.22	0	0	1	0,22	0,780	0,220
0.25-0.5	22	0.22	3	0.03	0,97	0,44	0,561	0,217
0.5-0.75	14	0.14	9	0.09	0,88	0,58	0,437	0,130
					0,73	0,77	0,355	0,153
					0,62	0,88	0,398	0,072
					0,47	0,95	0,532	0,038
					0,31	1	0,690	0,020
					0,21	1	0,790	0
					0,1	1	0,900	0
					0,05	1	0,950	0
					0,02	1	0,980	0
					0	1	1,000	0
					0	1	1,000	0
$AUC =$								0.85



Пример из радиационной генетики:



Лаборатория экологической генетики ИОГен РАН

Частоты хромосомных aberrаций в зависимости от генотипов по кандидатным генам у ликвидаторов аварии на ЧАЭС и в контроле



Контроль :
98 курсантов



Генотипирование по 20 генам
репарации, оксидативного стресса и
детоксикации ксенобиотиков



Ликвидаторы :
84 человека



Рутинный
цитогенетический
анализ

Частоты хромосомных aberrаций в зависимости от генотипов по кандидатным генам у ликвидаторов аварии на ЧАЭС и в контроле

**Рисковый генотип:
 минорные аллели в сайтах *XPD* и функциональные аллели генов *GST***



Контроль :
98 курсантов

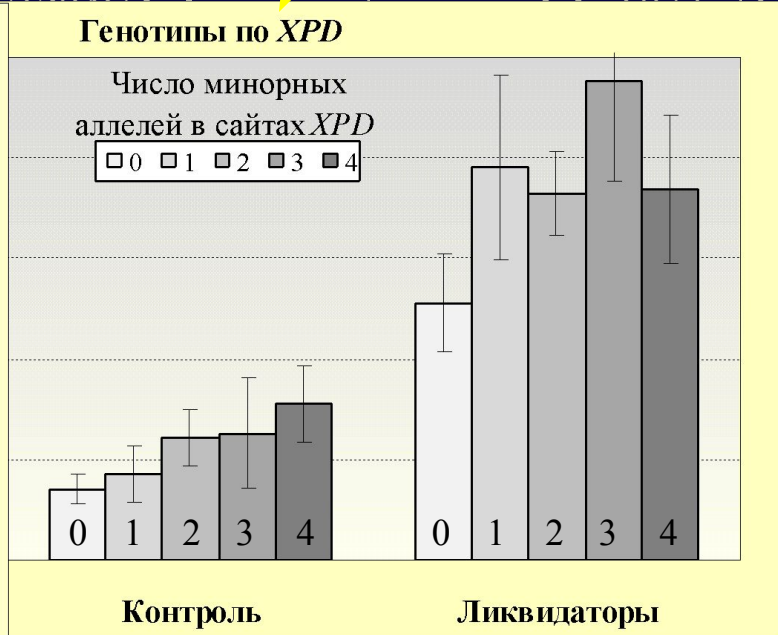
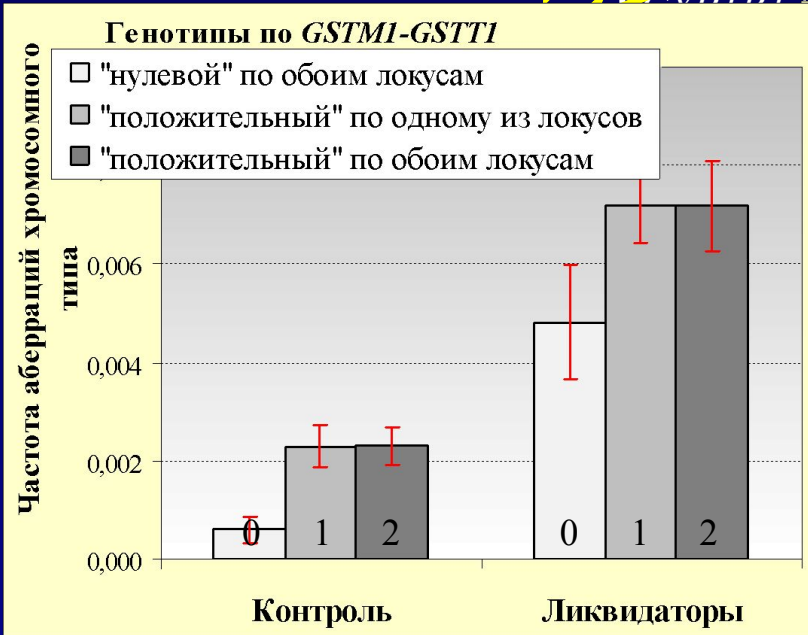


Ликвидаторы :
84 человека

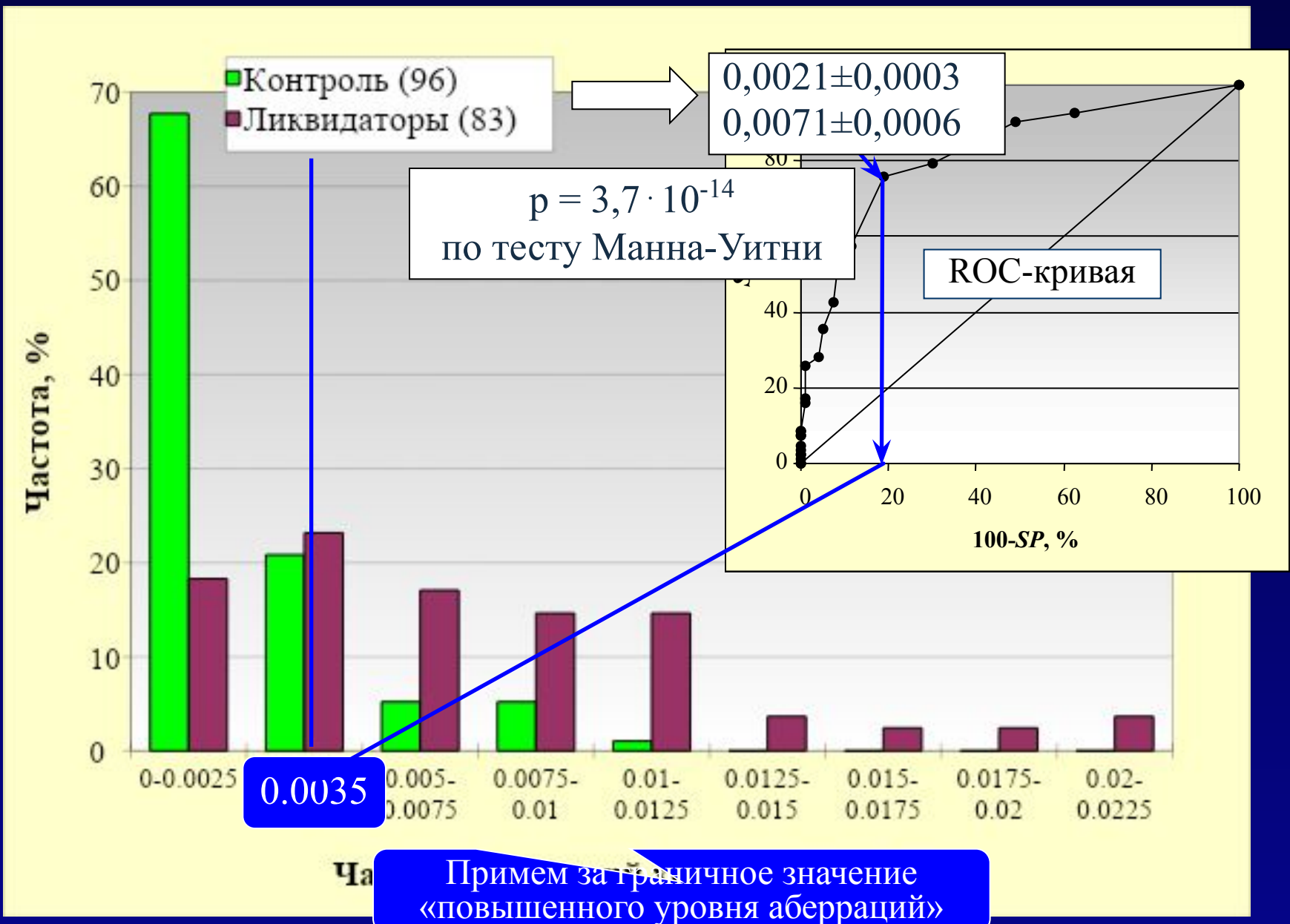
Ген	Сайты
<i>GSTM1</i>	Ins/Del
<i>GSTT1</i>	Ins/Del
	T3801C
<i>CYP1A1</i>	T606G
	A4889G

Ген	Сайты
<i>XRCC1</i>	C589T G1996A
<i>XPD (ERCC2)</i>	T2251G G862A
	G262T
<i>ERCC1</i>	T354C
<i>APTX1</i>	T444G

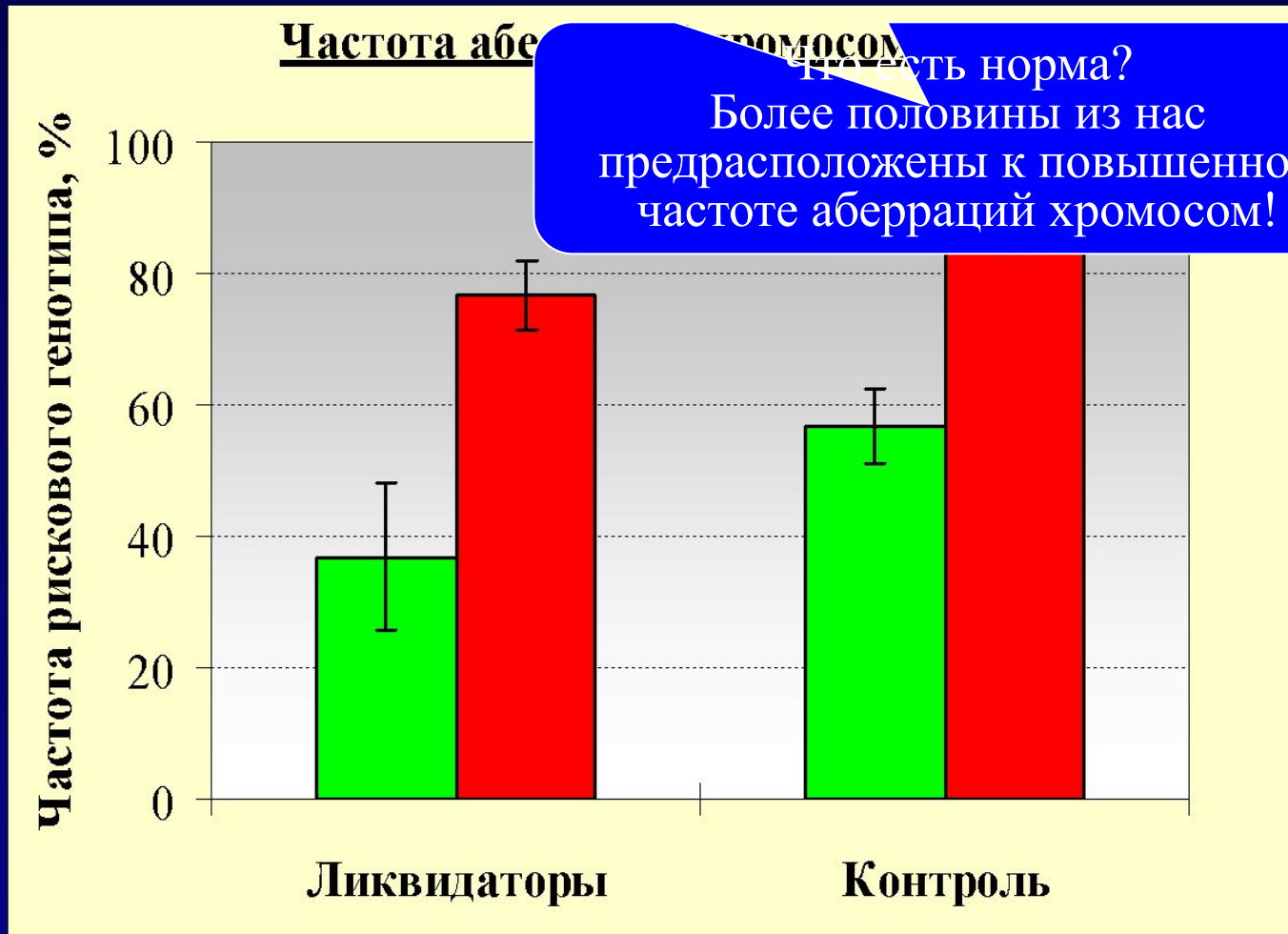
Ген	Сайты
<i>SOD2</i>	C47T
<i>CAT</i>	T21A
<i>GCLC</i>	C129T
<i>MTHFR</i>	C677T
<i>Tp53</i>	G215C



Что можно считать повышенным уровнем хромосомных aberrаций?



	Частота рискового генотипа, %	<i>SE</i> , %	<i>SP</i> , %	<i>AUC</i>	<i>OR</i>	<i>P</i>
Контроль	63.5	90.0	43.4	0.67	6.9	0.008
Ликвидаторы	67.5	76.6	63.2	0.70	5.6	0.002



и еще раз напомним - не ORом единым!

- ❑ Необходимо оценивать, чувствительность, относительный риск и прогностическую эффективность маркера (вычислять AUC)
- ❑ Хороший классификатор – это когда $OR > 5$, а частота маркера не менее 30%
- ❑ При $OR < 2.5$ маркер является заведомо плохим классификатором ($AUC < 0.6$ при любых частотах заболевания и маркера)
- ❑ Редкие мутации - малоэффективные классификаторы. Если частота маркера мала ($p_M < 0.2p_D$), то прогностическая эффективность теста крайне низка даже при $OR \gg 1$
- ❑ Не надо думать про OR , что относительный риск. Оцените RR как $OR / (1 - p_D + p_D \cdot OR)$
- ❑ Если вы используете маркер, частота встречаемости которого равна p_M , то чувствительность теста не превышает $p_M \cdot OR$, а специфичность теста скорее всего будет равна $1 - p_M$