

Исследование влияния биомаркеров на общую выживаемость пациентов с немелкоклеточным раком легкого (НМРЛ) при помощи обобщенных аддитивных моделей (GAM)

Кондрашина Анна, M&S Decisions

Косинский Юрий, M&S Decisions

01 ноября 2023

XV конференция
«Математические модели и
численные методы в
биологии и медицине»

Предпосылки работы

Немелкоклеточный рак легкого (НМРЛ) – распространенный тип рака, ассоциированный с низкой выживаемостью

Стадия TNM	Стадия	5-ти летняя выживаемость
I Стадия T1-T2, N0, M0	Локализованный	64%
II Стадия T2-T4, N0, M0	Региональный	37%
III - IV Стадии T1-T4, N1-N3, M0	Удаленный	8%
IV Стадия T1-T4, N1-N3, M1	Запущенный	3%

Предпосылки работы

- Успех таргетных методов лечения создал новую парадигму лечения пациентов с запущенным НМРЛ.
- Количественный анализ необходим для раннего выявления пациентов с хорошим и плохим прогнозом выживаемости, а также для оптимизации терапии и дизайна клинических исследований.
- Существующим методам такой аналитики не хватает способности к обобщению. Для ее увеличения необходимо использовать дополнительную информацию, например, данные о динамике биомаркеров.

Цель работы

Изучить влияние различных широко используемых в клинике лонгитюдных биомаркеров на функцию риска и выживаемости пациентов

Вопрос, на который хотелось бы ответить:

Могут ли лонгитюдные биомаркеры дать дополнительную информацию для прогнозирования выживаемости пациентов, недоступную при использовании только начальных показателей?

Данные

Название испытания	Описание контрольной руки	Количество пациентов	Биомаркеры	Количество измерений
INTEREST	<u>Docetaxel</u> , пациенты с НМРЛ, ранее проходившие терапию	520	ALT, ALP, AST, CREAT, NEUT, SLD, WBC	26370
IPASS	<u>carboplatin/paclitaxel</u> , первая линия терапии НМРЛ	606	ALT, ALP, AST, CREAT, NEUT, SLD, WBC	31503
ZD6747	<u>Erlotinib</u> , пациенты с НМРЛ, ранее проходившие терапию	569	ALT, ALP, AST, CREAT, LDH, NEUT, SLD, WBC	51405
ZODIAC	<u>Docetaxel</u> , вторая линия терапии НМРЛ	512	ALT, ALP, AST, CREAT, LDH, NEUT, NLR, SLD, WBC	33589

Ограничение модели пропорциональных рисков

Определение функции риска: $\lambda(t) := \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$

Модель пропорциональных рисков:

$$\lambda_i(t | \mathbf{x}_i) = \lambda_0(t) \exp(\mathbf{x}_i^T \boldsymbol{\beta}), \quad i = 1, \dots, n,$$

Почему
“пропорциональных”?

Для двух пациентов, i-го и j-го

$$\frac{\lambda(t | \mathbf{x}_i)}{\lambda(t | \mathbf{x}_j)} = \frac{\lambda_0(t) \exp(\mathbf{x}_i^T \boldsymbol{\beta})}{\lambda_0(t) \exp(\mathbf{x}_j^T \boldsymbol{\beta})} = \exp((\mathbf{x}_i^T - \mathbf{x}_j^T) \boldsymbol{\beta}) = \text{const}(t)$$

Клинические исследования показывают,
что это предположение не всегда верно

Кусочно-экспоненциальные аддитивные модели

Кусочно-экспоненциальная модель

$$\lambda_i(t|\mathbf{x}_i) = \lambda_j \exp \left(\sum_{k=1}^p x_{ij}^k \beta_k \right), \quad \forall t \in (t_{j-1}, t_j]$$

$$\lambda_i(t|\mathbf{x}_i) = \exp \left(\log(\lambda_j) + \sum_{k=1}^p x_{ij}^k \beta_k \right), \quad \forall t \in (t_{j-1}, t_j]$$

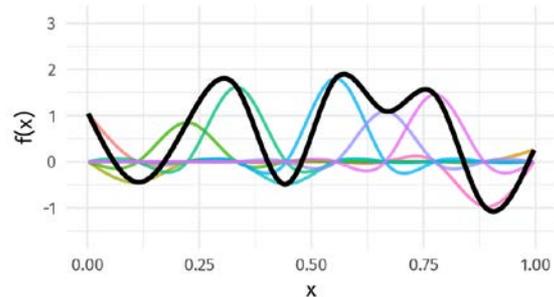
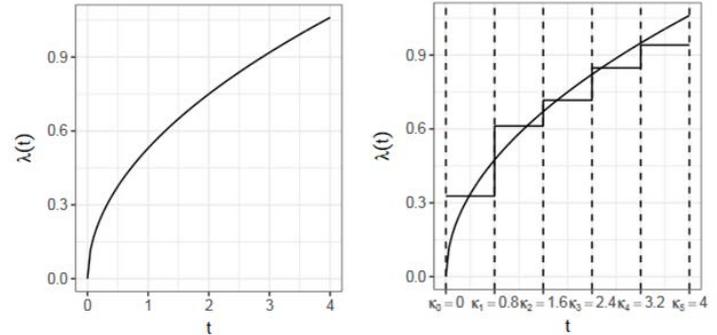
Обобщенная аддитивная модель

$$y = \sum_{k=1}^p f_k(x_k) \quad f_k(x_k) = \sum_{l=1}^m \beta_l b_l(x_k)$$

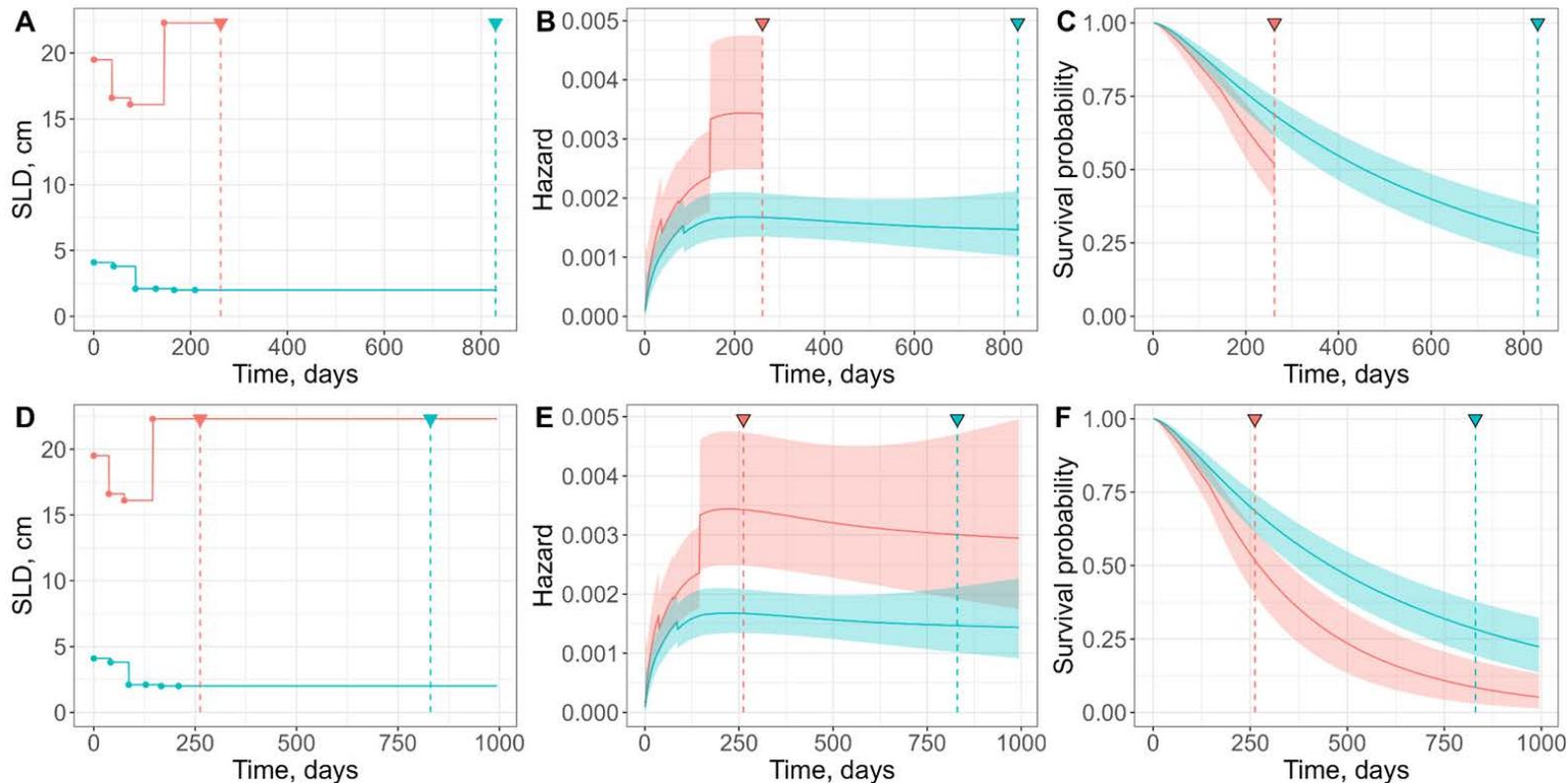
Обобщенная аддитивная + Кусочно-экспоненциальная
=

Кусочно-экспоненциальная аддитивная модель:

$$\lambda(t|\mathbf{x}_i) = \exp \left(f_0(t) + \sum_{k_1=1}^{p_1} f_{k_1}(x_{ij}^{k_1}) + \sum_{k_2=1}^{p_2} \beta_{k_2} x_i^{k_2} \right)$$



Пример оценки функций выживаемости и риска индивидуальных пациентов



Метрика ROC-AUC

Задача бинарной классификации:

$$X \in \mathbb{R}^n, Y \in \{0, 1\} \quad Y_{pred} = \begin{cases} 1, f(X) \geq c \\ 0, f(X) < c \end{cases}$$

Доля истинно положительных (TPR):

$$TPR(c) = \frac{TP}{AIPP} = \frac{TP}{TP + FN} = P(Y = 1 | f(X) \geq c)$$

Доля ложно положительных (FPR):

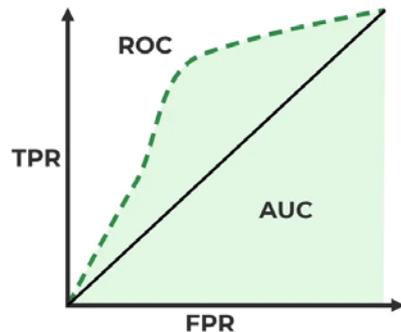
$$FPR(c) = \frac{FP}{AIIIN} = \frac{FP}{TN + FP} = 1 - P(Y = 0 | f(X) < c)$$

ROC кривая отображает соответствие TPR и FPR.

Чем выше кривая – тем лучше

ROC-AUC – площадь под ROC кривой

	$Y == 1$	$Y == 0$
$Y_{pred} == 1$	Истинно положительные (TP)	Ложно положительные (FP)
$Y_{pred} == 0$	Ложно отрицательные (FN)	Истинно отрицательные (TN)



Метрика ROC-AUC позволяет оценивать насколько хорошо модель различает классы и сравнивать модели между собой

Метрика ROC-AUC в анализе выживаемости

В каждый момент времени исследования можем разделить всех пациентов на тех, кто уже испытал событие и тех, кто еще не испытал.

$$Y_i = \mathbb{I}(t < T_i)$$

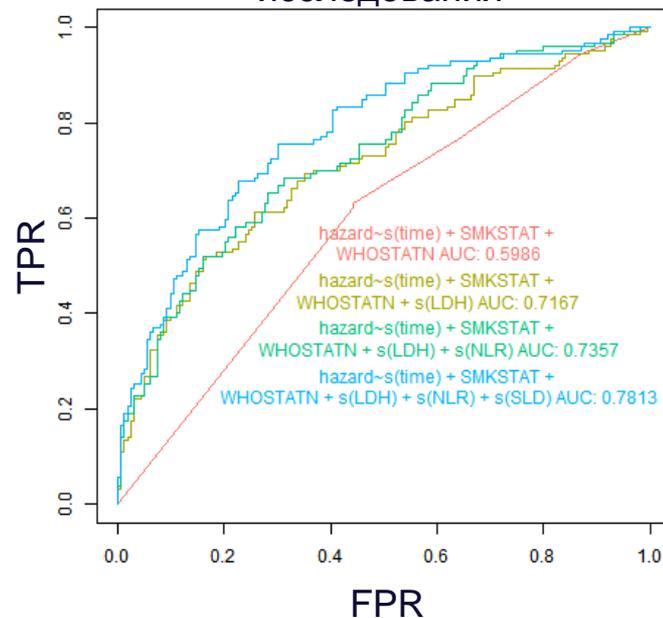
Таже можно вычислить функцию выживаемости в каждый момент времени и на ее основе классифицировать пациентов.

$$S_i(t) = P(T_i > t)$$

По этим данным можно посчитать TPR и FPR, построить ROC кривую и высчитать ROC-AUC

$$FPR(t, c) = 1 - P(0|S(t) > c) \quad TPR(t, c) = P(1|S(t) < c)$$

ROC кривые на 300-ый день исследования



Полученные зависимости ROC-AUC от времени можно использовать для сравнения моделей

Метрика Brier Score (BS) в анализе выживаемости

В задаче бинарной классификации функция $f(x)$ может быть интерпретирована как **вероятность** принадлежности объекта к положительному классу.

$$X \in \mathbb{R}^n, Y \in \{0, 1\} \quad Y_{pred} = \begin{cases} 1, & f(X) \geq c \\ 0, & f(X) < c \end{cases}$$

Метрика Brier Score (BS) используется для оценки **точности** предсказанных вероятностей

$$BS = \frac{1}{N} \sum_{i=1}^N (f(X_i) - Y_i)^2$$

В анализе выживаемости BS может быть вычислен в каждый момент времени исследования

$$BS(t) = \frac{1}{N} \sum_{i=1}^N ((1 - S_i(t)) - Y_i)^2 = \frac{1}{N} \sum_{i=1}^N (\tilde{Y}_i - S_i(t))^2$$

Временные зависимости BS можно использовать для сравнения моделей по точности предсказаний

Алгоритм отбора ковариат



Построение наилучшей модели

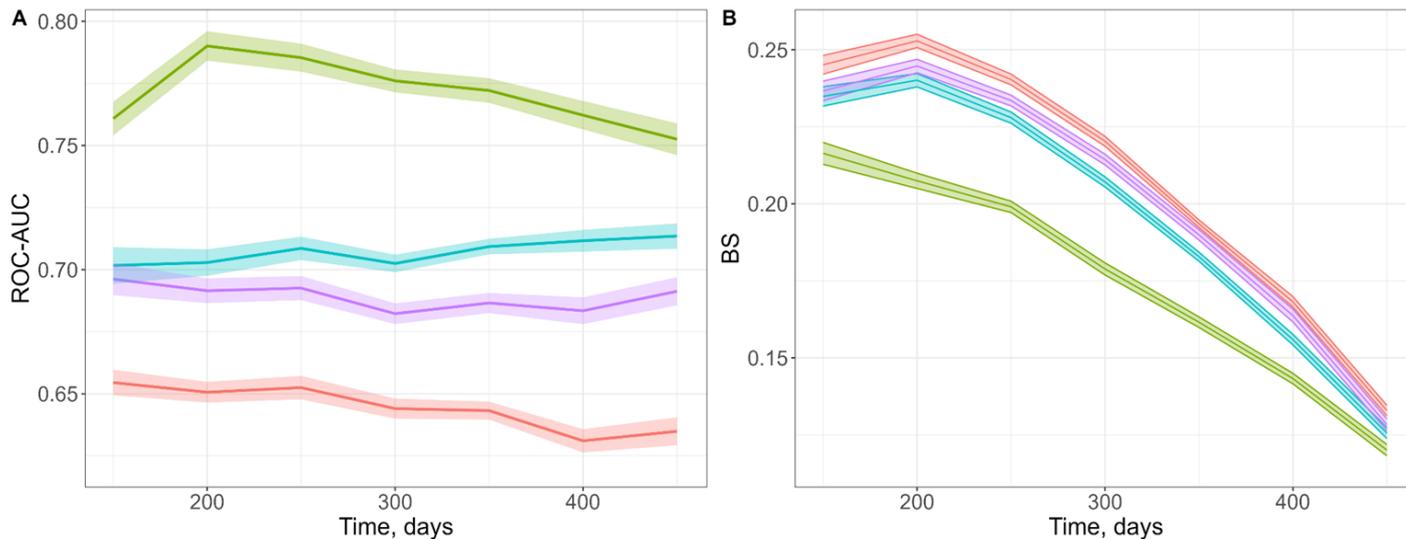
$$\text{Hazard} \sim \text{Базовый риск } s(\text{time}) + \text{Базовые категориальные ковариаты } \text{WHOSTAT} + \text{SMKSTAT} + \text{Эффект от лонгитудальных ковариат } \text{SLD} + s(\text{ALP}) + s(\text{CREAT})$$

Основные этапы отбора ковариат для лучшей модели для исследования INTERET

С л о ж н о с т ь ↓	Модель	AIC	О п и с а т е л ь н о с т ь ↓		
			p.val of SLD	p.val of s(ALP)	p.val of s(CREAT)
	Hazard~s(time)	3208.783			
	+ SMKSTAT + WHOSTATN	3185.521			
	+ SLD _b	3172.476			
	+ SLD	3158.894	1.19e-08		
+ s(ALP) + s(CREAT)	3113.062	1.03e-06	< 2e-16	1.52e-05	

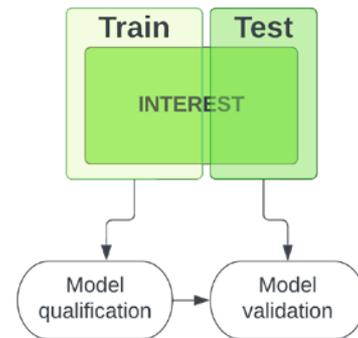
Валидация модели на данных одного испытания при помощи метрик ROC-AUC и BS

Средние значения метрик ROC-AUC и BS с 95% доверительными интервалами для различных моделей исследования INTEREST



Legend for models:

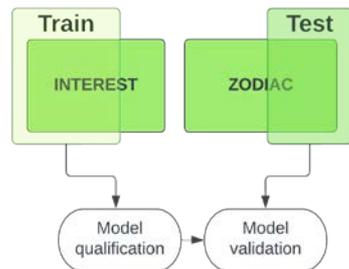
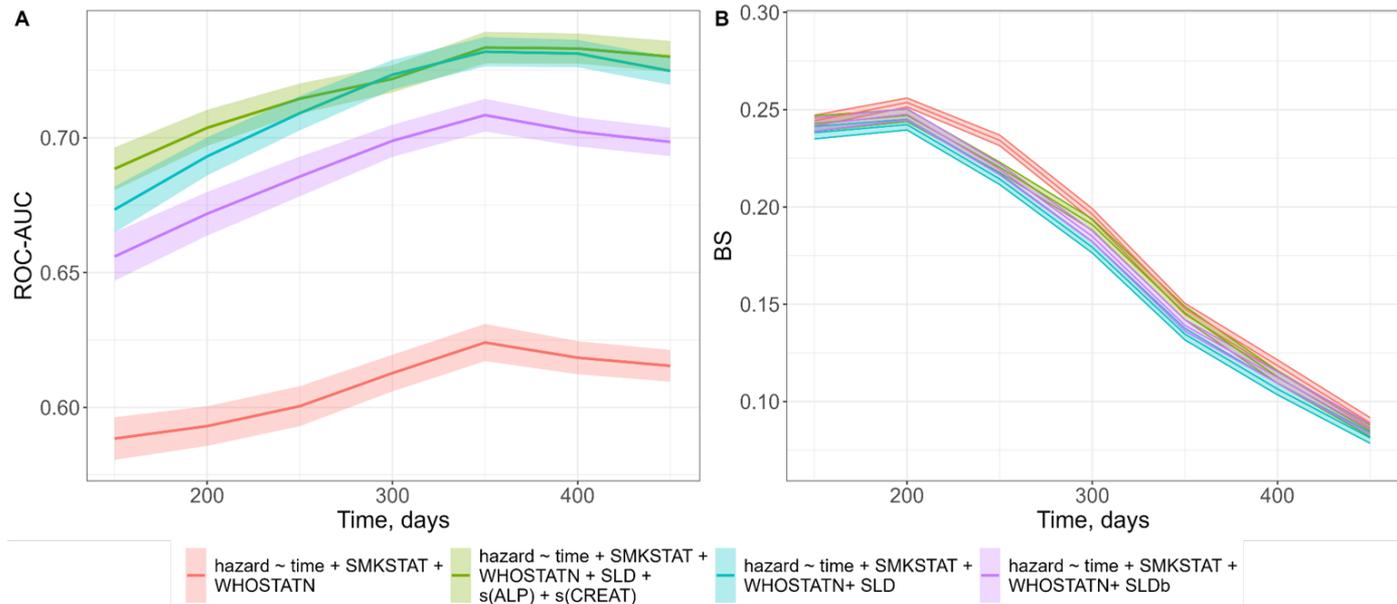
- Red: hazard ~ time + SMKSTAT + WHOSTATN
- Green: hazard ~ time + SMKSTAT + WHOSTATN + SLD + s(ALP) + s(CREAT)
- Cyan: hazard ~ time + SMKSTAT + WHOSTATN + SLD
- Purple: hazard ~ time + SMKSTAT + WHOSTATN + SLDb



Модели, учитывающие данные лонгитудинальных биомаркеров, имеет лучшую предиктивную способность, чем модели, учитывающие только исходную информацию

Валидация на данных из другого КИ

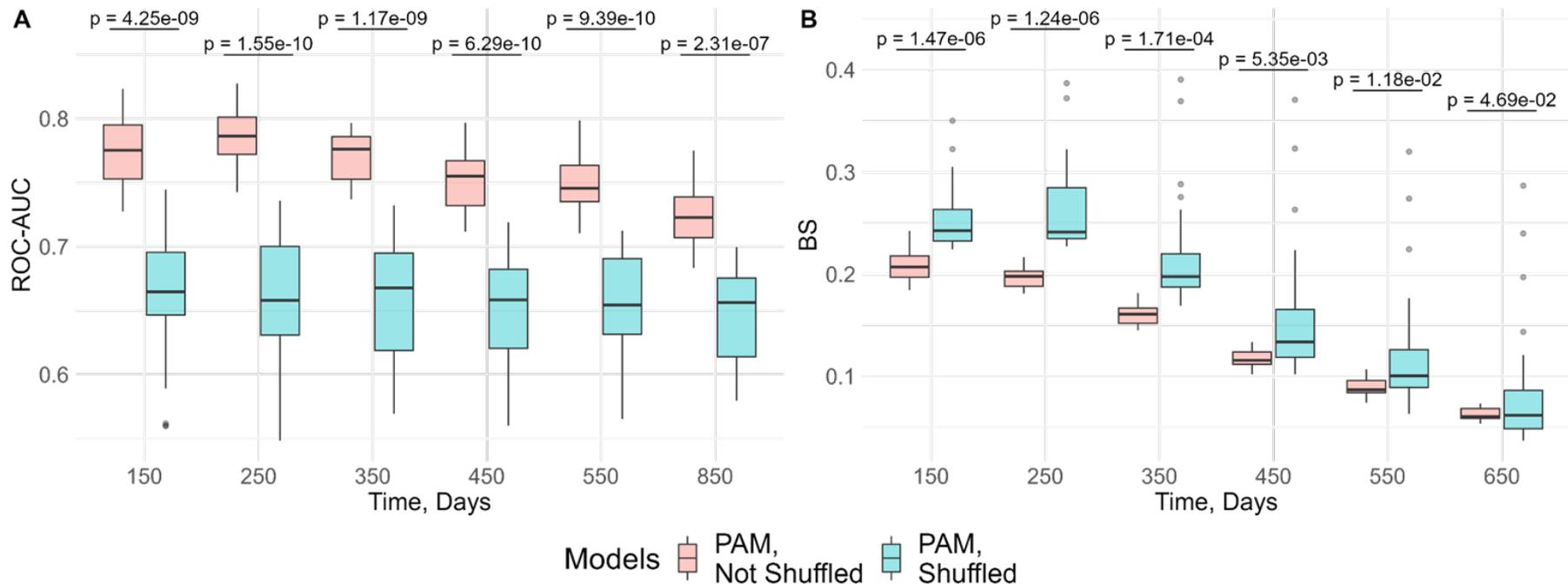
Средние значения метрик ROC-AUC и BS с 95% доверительными интервалами для различных моделей исследования INTEREST. Метрики вычислены по данным из ZODIAC



Предсказательная способность всех моделей ухудшилась, но, как и прежде,

15 у моделей с лонгитюдными биомаркерами лучшая предиктивная способность

Перестановочный тест



Добавление ковариат, зависящих от времени, улучшает показатели ROC-

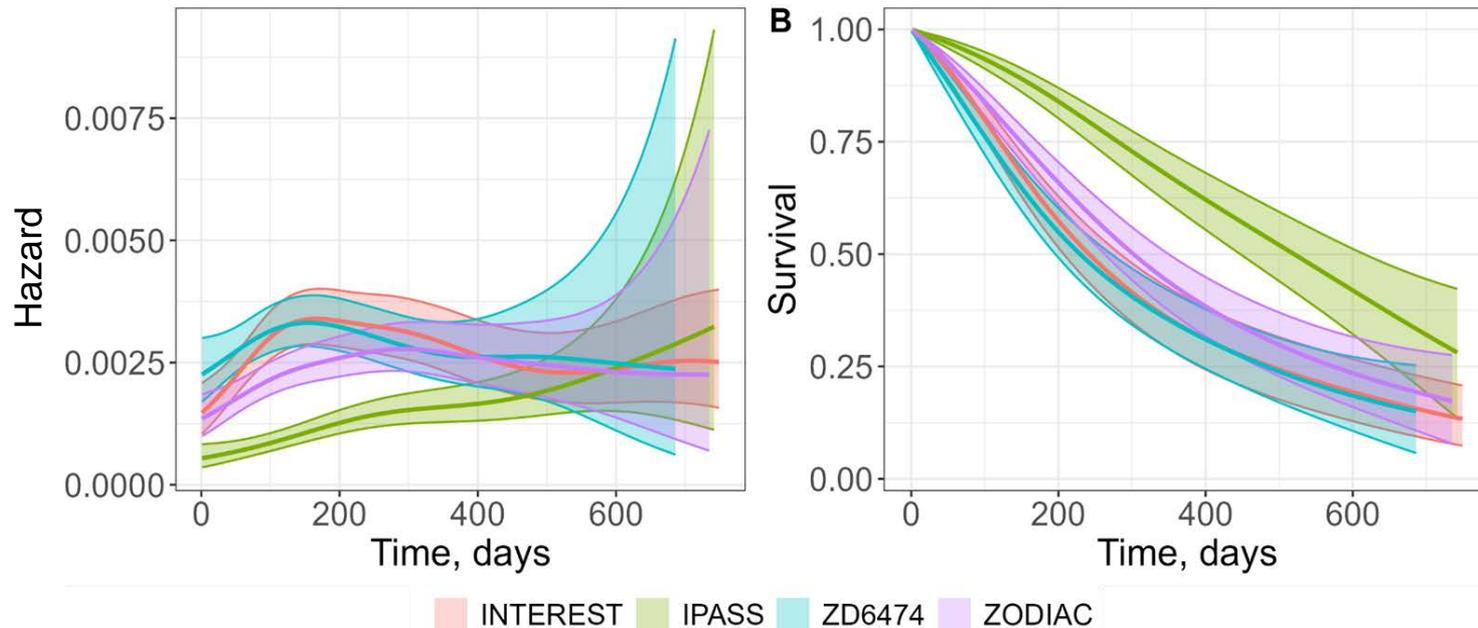
AUC и BS

Важные ковариаты для лучших моделей из разных исследований

Исследование	ALP	CREAT	LDH	NEUT	NLR	SLD
INTEREST	+	+	X	-	X	+
IPASS	-	-	X	+	X	+
ZD6474	-	-	+	-	+	+
ZODIAC	-	-	+	-	+	+

- + Биомаркер был важен для лучшей модели
- Биомаркер не был важен для лучшей модели
- x Биомаркер не измерялся в исследовании

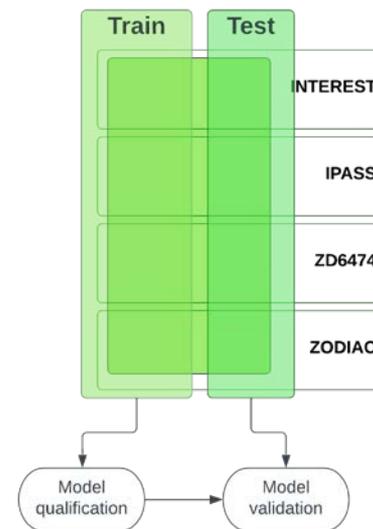
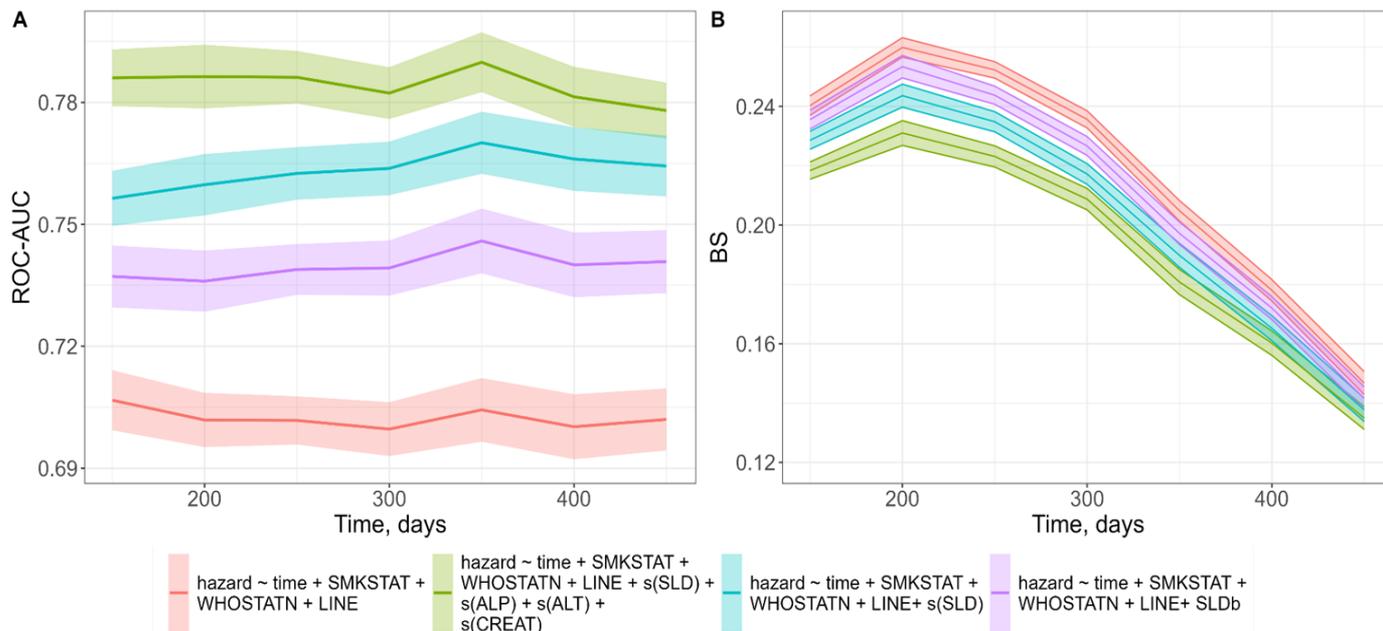
Базовые функции риска и выживания для общей модели выживания



Базовые функции риска и выживания для набора данных IPASS отличаются от функций для других трех наборов данных. IPASS было единственным исследованием терапии первой линии. Следовательно, линия терапии должна быть включена в окончательную модель как ковариата

Общая модель выживаемости на основе четырех наборов данных

Средние значения метрик ROC-AUC и BS с 95% доверительными интервалами для различных моделей всех исследований



Модели, использующие данные лонгитудинальных биомаркеров, имеют лучшую предиктивную способность.

Выводы

- Исследовали значение лонгитюдных биомаркеров ALP, ALT, AST, CREAT, LDH, NLR, NEUT и SLD для прогнозирования выживаемости на основе четырех клинических исследований.
- Было продемонстрировано прогностическое влияние категориальных ковариат «Статус курения», «Общее состояние» и исходный размер опухоли (SLDb) на дифференциацию пациентов по выживаемости.
- Модели, учитывающие данные о лонгитудинальных биомаркерах, имеют лучшую прогностическую способность, чем модели, учитывающие только исходную информацию.
- Важным для прогнозов моделей из всех четырех исследований и общей модели лонгитудинальным биомаркером стал размер опухоли (SLD).
- Методология, используемая в этом исследовании, помогает выбрать релевантные биомаркеры, которые наряду с SLD могут быть использованы для построения семи-механистических моделей динамики этих биомаркеров и их влияния на функцию риска.