



Modeling & Simulation  
Decisions

# Совместное моделирование измеряемых во времени биомаркеров и выживаемости при онкологических заболеваниях: статистические методы и графическая диагностика

С.Гаврилов, К.Жуденков, А.Софронова, О.Степанов, К.Песков (M&S Decisions)

02 Ноября 2020

Доклад на конференции

# Joint Modeling (Совместное моделирование)

- Главная задача **Joint Modeling (JM)**: установление связи между лонгитюдными (измеряемыми во времени) биомаркерами и клинически значимыми событиями.
- Связь лонгитюдных биомаркеров и событий выражена посредством **функции риска (hazard)**, структура которой обобщает модель пропорциональных рисков Кокса.
- Для описания лонгитюдных биомаркеров используются модели смешанных эффектов.
- Параметры модели определяются из совместной функции правдоподобия, которая объединяет функции правдоподобия модели смешанных эффектов и событийной составляющей.

Vector of Individual baseline covariates \* Vector of corresponding coefficients. **Similar to COX proportional hazard.**

$$\text{Individual\_Hazard}(t) = \text{Baseline\_Hazard}(t) * \exp [ (\boldsymbol{Y}, \boldsymbol{\omega}) + f(\boldsymbol{\alpha}_{slid}, \mathbf{SLD}(t)) ]$$

Impact of individual multivariate longitudinal history processed by Mixed Effect model. **Extension to COX.**

# Общая характеристика Совместного моделирования

- Преимущества:
  - Явная структура модели, связывающая значения биомаркеров и вклад в функцию риска;
  - Включает модель смешанных эффектов для лонгитюдных биомаркеров, что является общепризнанным способ их описания;
  - После определения параметров модели она может быть использована для предсказания вероятности событий и значений биомаркеров на основе индивидуальных лонгитюдных данных;
- Ограничения подхода:
  - Требует затратных с вычислительной точки зрения процедур для определения параметров и выполнения предсказаний;
  - Возможность определения оптимальных параметров модели зависит от многих факторов;
  - Использование модели для предсказаний предполагает генерацию выборки из апостериорного распределения параметров, что может представлять сложность;

# Предназначение Совместного моделирования

- Описание поведения биомаркеров. Благодаря использованию совместной функции правдоподобия коэффициенты модели смешанных эффектов подвержены влиянию событийной составляющей (и наоборот), что может быть интерпретировано как более полноценное описание лонгитюдных биомаркеров.
- Использование модели для предсказаний. Возможность Совместных моделей давать предсказания может быть эффективно использована для персонафицированного терапевтического подхода и анализа клинических исследований.

# Использование JM в проектах M&S Decisions

- **Анализ выживаемости для немелкоклеточного рака легких** (клинические исследования ATLANTIC, 1108 (*durvalumab*), IPASS (*gefitinib*))
  - S Gavrilo et al. Longitudinal assessment of tumor size and neutrophil count in multivariate joint models are more predictive of survival than their baseline values in patients with non-small cell lung cancer. **PAGE Meeting**, 2019, Stockholm, Sweden.
  - K. Zhudnikov, S Gavrilo et al. Longitudinal tumor size and NLR are more predictive of individual survival than their baseline values in patients with non-small cell lung cancer treated with durvalumab. 2019 **ASCO Annual Meeting**, Chicago, USA.
- **Анализ выживаемости метастатического рака предстательной железы** (на основе данных ProjectDataSphere)
  - S Gavrilo et al. Evaluation of longitudinal biomarkers associated with survival of prostate cancer patients by means of Bayesian joint modeling. **Stancon 2019**, Cambridge, UK

# Совместная функция правдоподобия

Вклад событийной составляющей

$$\mathcal{S}(t) = \exp\{-\mathcal{H}(t)\} = \exp\left\{-\int_0^t h(s) ds\right\} \quad (1)$$

$$\ell(\theta) = \sum_{i=1}^n \delta_i \log p(T_i; \theta) + (1 - \delta_i) \log \mathcal{S}_i(T_i; \theta) = \sum_{i=1}^n \delta_i \log h_i(T_i; \theta) - \int_0^{T_i} h_i(s; \theta) ds \quad (2)$$

Вклад модели смешанных эффектов

$$p(y_i | b_i; \theta) = \prod_{j=1}^{n_i} p(y_{ij} | b_i; \theta) \quad (3)$$

$$\ell(\theta) = \sum_{i=1}^n \log p(y_i; \theta) \quad (4)$$

**log (Совместная функция  
правдоподобия) = (2) + (4)**

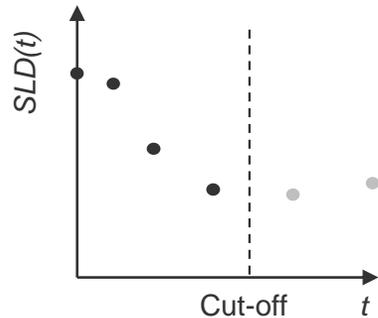
# Принцип предсказаний на основе Совместной модели

Накопление индивидуальной лонгитюдной информации вплоть до Cut-off времени

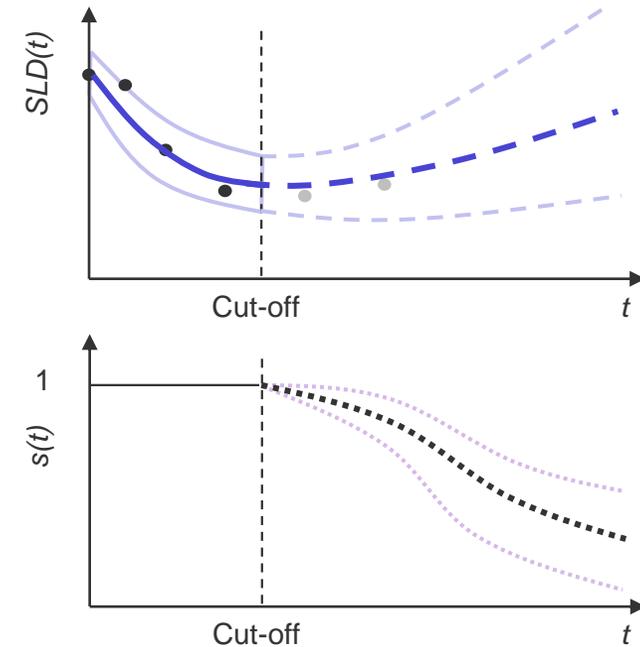
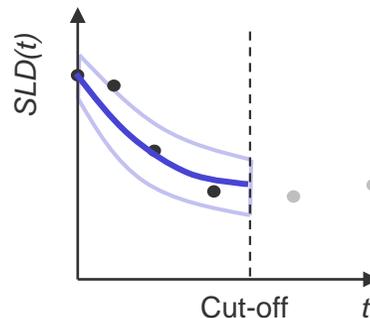
Расчет индивидуальных коэффициентов в модели смешанных эффектов

Предсказания: вероятность выживания  $s(t)$  и предсказания лонгитюдного биомаркера  $SLD(t)$  для  $t > \text{Cut-off}$

Известная лонгитюдная информация вплоть до времени Cut-off:  
 $y_i = SLD(t_i)$



$SLD(t) = c \cdot \exp(d \cdot t) + b \cdot t$   
 $\{c, d, b\}$  вычисляется из апостериорного распределения  $pr(c, b, d | y_i)$



$$s(t) = e^{-\int_0^t h(t) dt}$$

# Некоторые программные средства для Совместного моделирования

	<b>JM package</b>	<b>JMbayes package</b>	<b>Monolix</b>	<b>NONMEM</b>	<b>Rstanarm (stanjm) package</b>	<b>STAN</b>
<b>R interface</b>	Full	Full	Command interface	Command interface	Full*	Command interface
<b>Dataset</b>	Time-to-event	Time-to-event	Vertical	Vertical	Time-to-event	Custom
<b>Nonlinear mixed-effect models</b>	No	No	Yes	Yes	No	Yes
<b>Multivariate models</b>	No	Yes	Yes	Yes	Yes**	Yes
<b>Splines</b>	Yes	Yes	Coding	Coding	Yes	Coding
<b>Predictions</b>	Implemented	Implemented	Only random effects	partially implemented	Implemented	Coding
<b>Overall Complexity</b>	low	low	medium	medium	low	high
<b>Modifications after fitting</b>	No	No	Yes	Yes	No	Yes

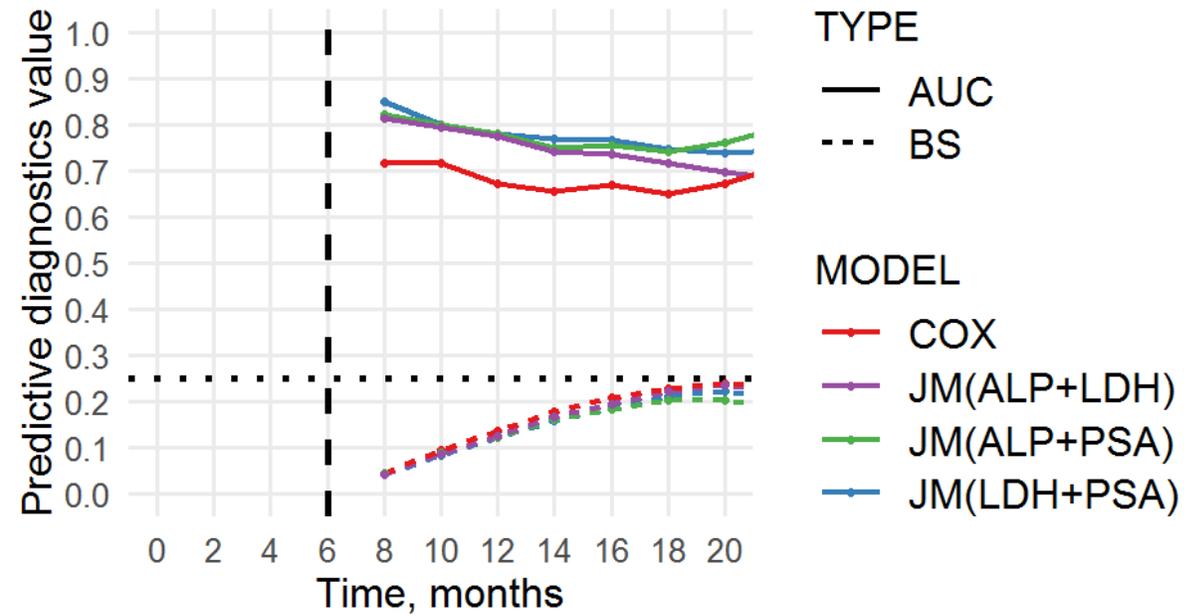
# Пакеты для R реализующие линейные JM

JM package	JMbayes package	Rstanarm package (stanjm)
<p>Pros</p> <ul style="list-style-type: none"><li>• Easy syntax</li><li>• <b>Robust and very fast fit</b></li><li>• Fully documented (Rizopoulos book)</li><li>• Various baseline hazard models possible</li><li>• complex association structure (derivative, integral, etc.) possible</li><li>• Coded in R</li></ul>	<p>Pros</p> <ul style="list-style-type: none"><li>• Easy syntax</li><li>• Multivariate capability</li><li>• Robust and rather fast fit</li><li>• Bayesian algorithm, priors possible</li><li>• Various baseline hazard models possible</li><li>• complex association structure (derivative, integral, etc.) possible</li></ul>	<p>Pros</p> <ul style="list-style-type: none"><li>• Easy syntax</li><li>• Multivariate capability*(see cons)</li><li>• Bayesian algorithm, priors possible</li><li>• Various baseline hazard models possible</li><li>• complex association structure (derivative, integral, etc.) possible</li></ul>
<p>Cons</p> <ul style="list-style-type: none"><li>• No multivariate models</li><li>• Non-Bayesian</li><li>• No access to random effects or modification of fitted model</li></ul>	<p>Cons</p> <ul style="list-style-type: none"><li>• <b>!! sequential, not honest joint model</b></li><li>• No access to random effects or modification of fitted model</li><li>• Limited prediction functionality</li></ul>	<p>Cons</p> <ul style="list-style-type: none"><li>• <b>Fitting is not robust and slow</b></li><li>• Limited multivariate capability: predictions are not possible due to NA issue</li><li>• Limited access to random effects and no modification of fitted model</li></ul>

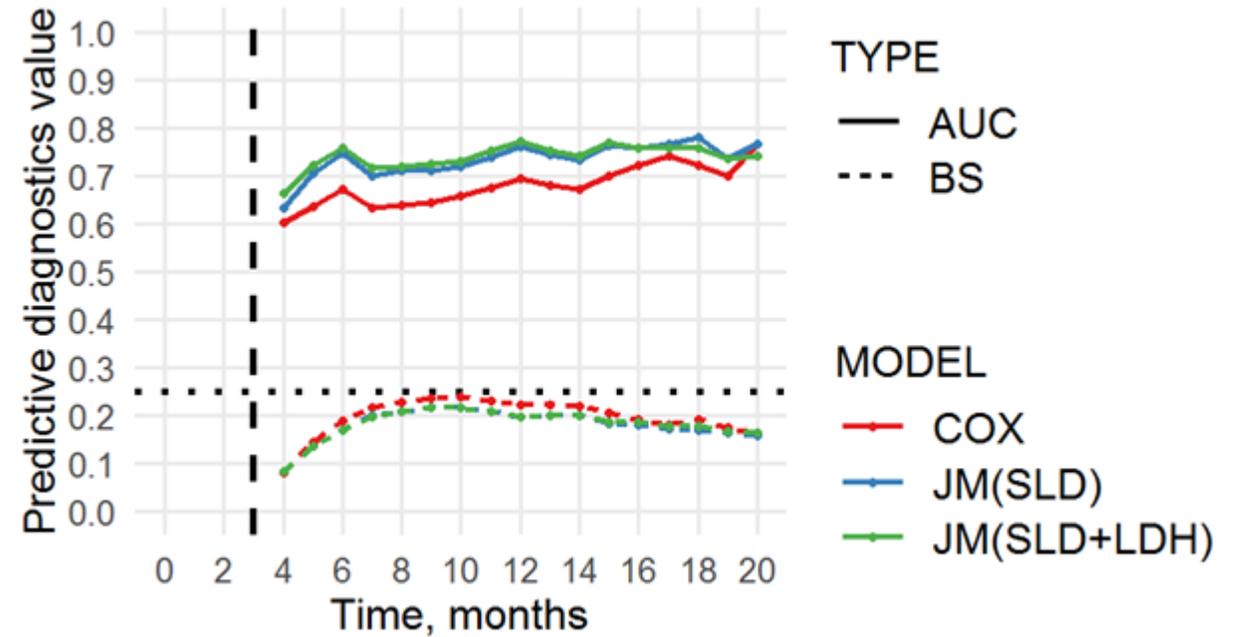
# Среды для создания нелинейных JM

Monolix	NONMEM	STAN
<p>Pros</p> <ul style="list-style-type: none"> <li>• <b>Differential equations solver</b></li> <li>• Easy covariate mechanism</li> <li>• Efficient SAEM algorithm</li> <li>• User-friendly graphical interface</li> <li>• Based on hazard definition</li> </ul>	<p>Pros</p> <ul style="list-style-type: none"> <li>• <b>Differential equations solver</b></li> <li>• Efficient fitting algorithms</li> <li>• Based on hazard definition</li> </ul>	<p>Pros</p> <ul style="list-style-type: none"> <li>• Fully customizable</li> <li>• Based on likelihood definition</li> <li>• Bayesian</li> </ul>
<p>Cons</p> <ul style="list-style-type: none"> <li>• Baseline hazard should be coded (no splines, etc.)</li> <li>• Slow computation</li> <li>• Time-consuming posterior diagnostics: except individual random effects estimation everything else should be coded in R</li> <li>• No access to likelihood function</li> </ul>	<p>Cons</p> <ul style="list-style-type: none"> <li>• Baseline hazard should be coded (no splines, etc.)</li> <li>• Time-consuming posterior diagnostics: requires some coding in R</li> <li>• Difficult to compose proper model</li> <li>• Access to likelihood function is strictly limited</li> </ul>	<p>Cons</p> <ul style="list-style-type: none"> <li>• <b>Coded from scratch</b></li> <li>• Difficult to achieve stable fitting (sampling)</li> </ul>

# Графическая диагностика на основе метрик ROC-AUC и Brier Score



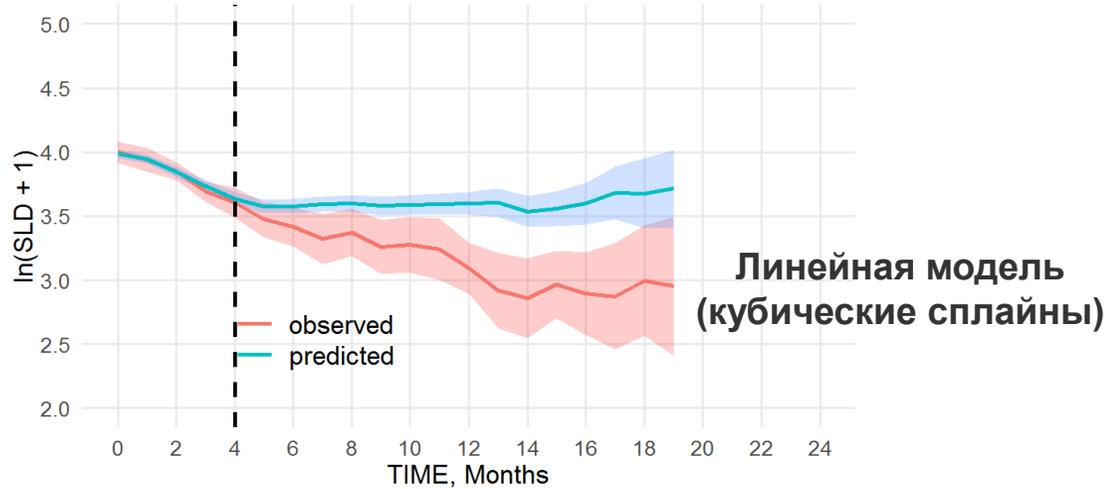
Пример 1. Графическая диагностика для сравнения точности предсказаний выживаемости различных моделей на основе данных клинических исследований рака предстательной железы.



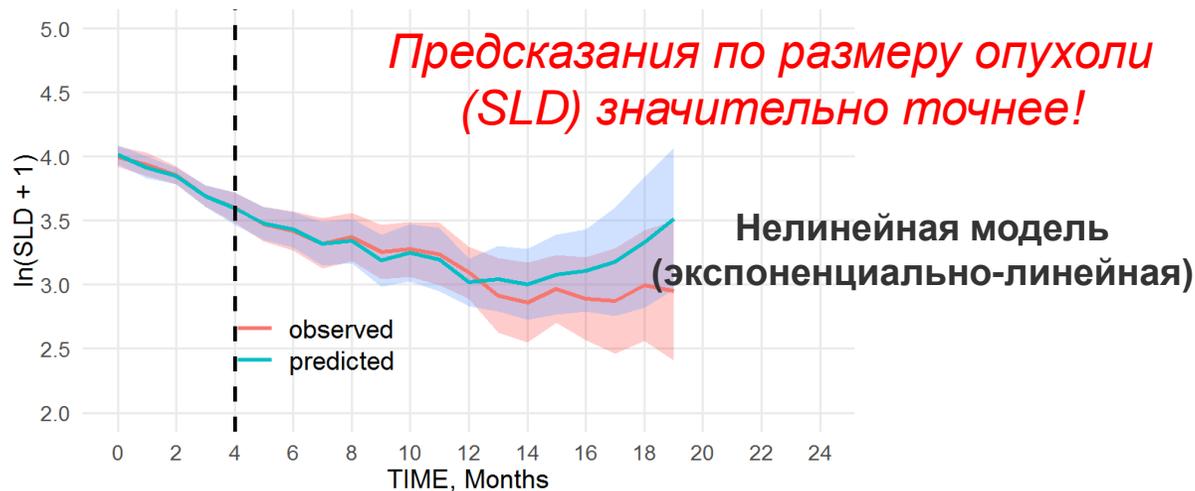
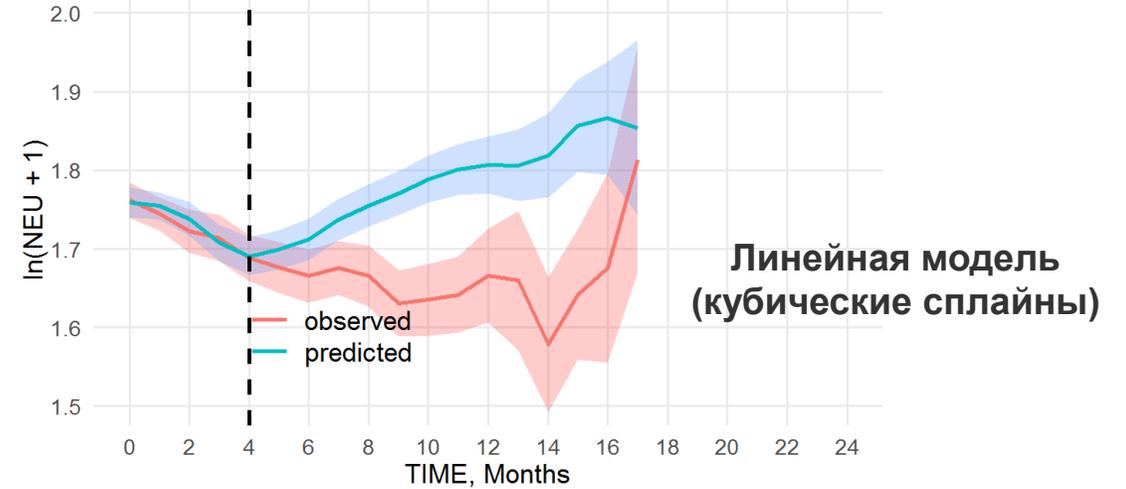
Пример 2. Графическая диагностика для сравнения точности предсказаний выживаемости различных моделей на основе данных клинических исследований немелкоклеточного рака легких.

# Сравнение точности лонгитюдных предсказаний линейных и нелинейных моделей

Популяционный профиль размера опухоли  $\ln(\text{SLD}+1)$



Популяционный профиль нейтрофилов  $\ln(\text{NEU}+1)$





Modeling & Simulation  
Decisions

**THANK YOU FOR ATTENTION**

