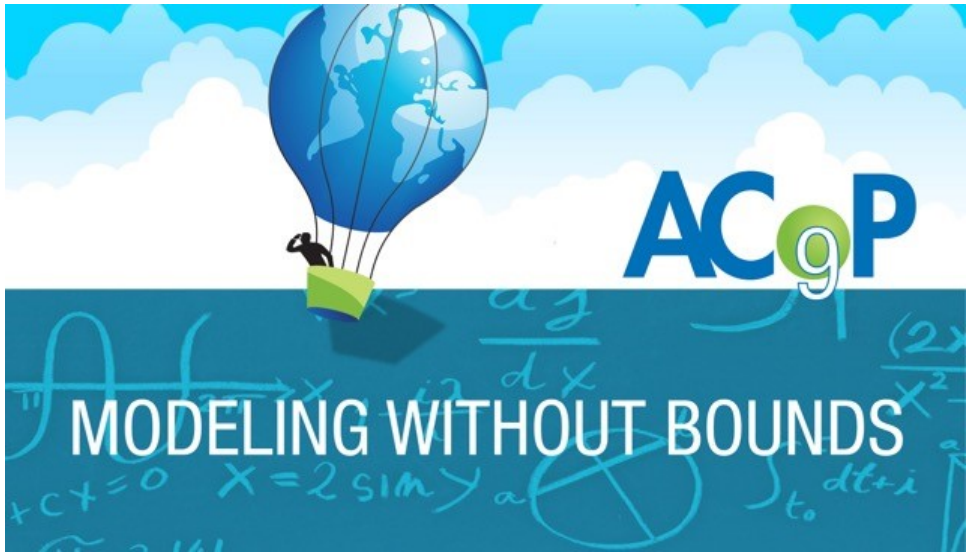


Session 3a:
Pharmacometrics in Big Data Era - Mission possible to find the needle in a haystack

Chairs:
Liang Zhao

Big Data Toolsets to Pharmacometrics: Application of Machine Learning for Time-to-Event Analysis

Meng Hu
Division of Quantitative Methods and Modeling
Office of Research and Standards
OGD | CDER | US FDA

Disclaimer

- The opinions expressed in this presentation are those of the speaker and may not reflect the position of the U. S. Food and Drug Administration

Background

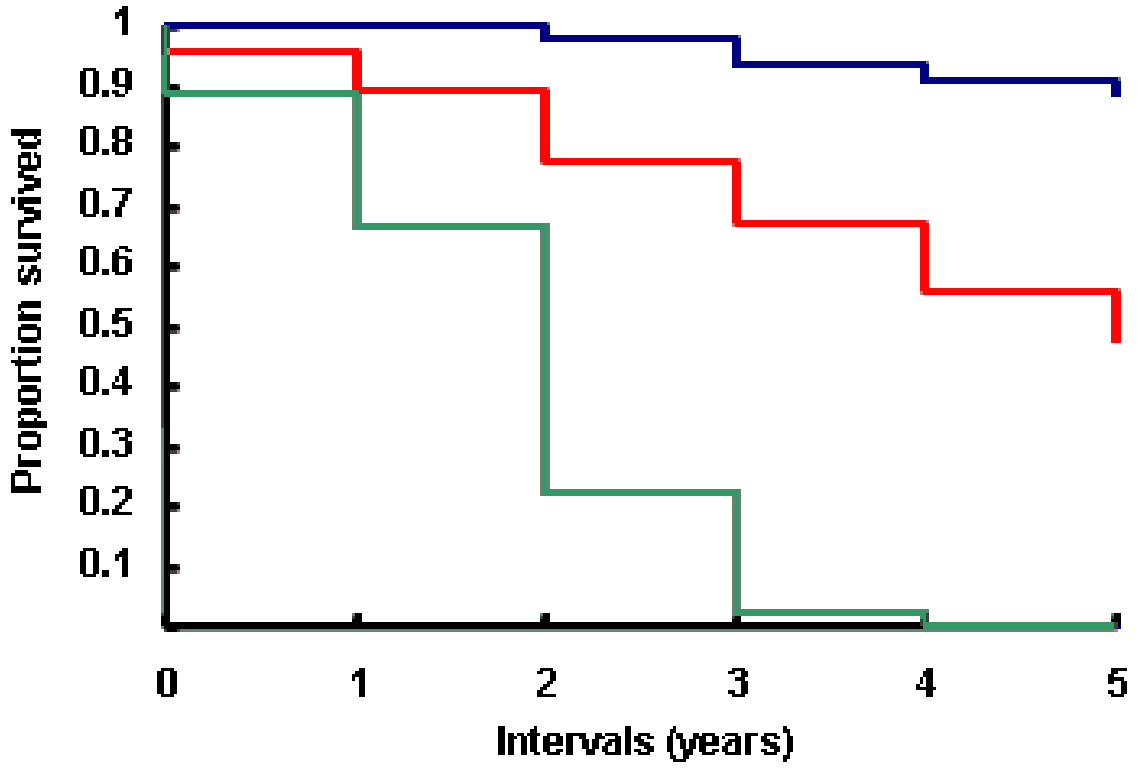
- In the Big Data era
  - 4V (Volume; Variety; Velocity; Veracity)

- Big Data toolsets
  - Artificial intelligence
  - Machine learning
  - Nature language processing
  - Others

- Can we take advantage of these tools?
  - **Application of Machine Learning for Time-to-Event Analysis**

Time-to-Event (Survival) Analysis

- To analyze the expected time to event occurrence (e.g., heart attack)



Time-to-Event (Survival) Analysis

- Heavy regulatory utilizations (e.g., clinical trial)
- Research opportunities (e.g., time to the ANDA submission)
- Featured by censoring data (i.e., no event during the study period)
- **Great needs on methodologies for the complex data** (e.g., nonlinear and high-dimension)

Conventional methods and challenges

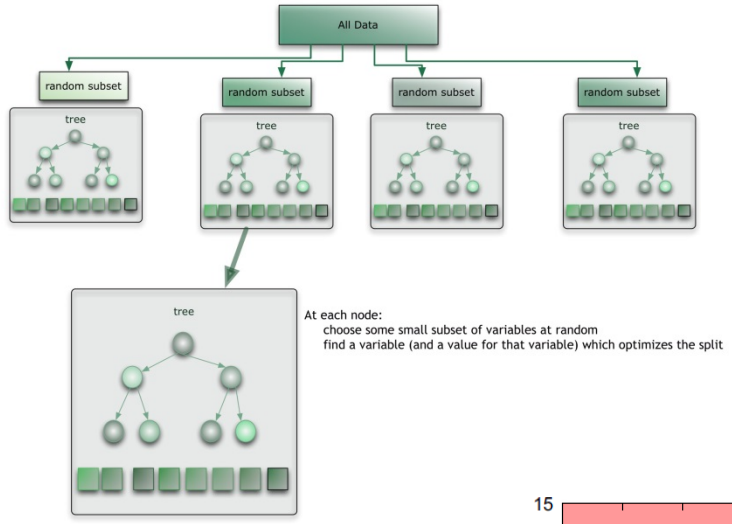
- Regression-based
  - Proportional hazard model (parametric)
  - Accelerated failure time model (parametric)
  - Cox proportional hazard model (semi-parametric)

$$h(t) = h_0(t) \times \exp(b_{age}.age + b_{sex}.sex + \ldots + b_{group}.group)$$

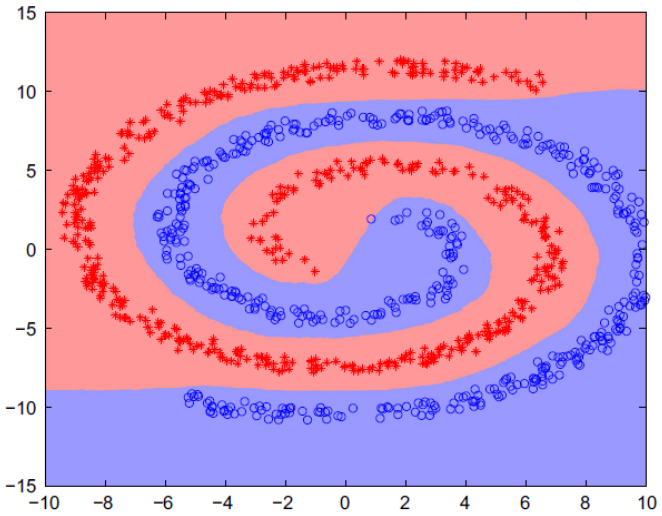
- Challenges
  - Predefined distribution for parametric methods
  - Proportional hazard function
  - Linear, additive relationships between predictors
  - Difficult to converge for high dimensional data

Machine Learning Methodologies for survival analysis

- Artificial neural network (ANN)
- Random survival forest (RSF)
- Support vector machine (SVM)
- Deep learning



- Advantages
  - Less distribution assumption
  - No assumption of linear relationship
  - No assumption of independency between predictors
  - Immune to large-feature problem



Aims of Research

- Systemically investigate the performances of the machine learning based methods and the conventional Cox model under various preset scenarios, e.g., with nonlinear predictors in the hazard function

- Check capabilities of the machine learning based methods for accommodating high-dimensional time-to-event data

Gong, et al. Clinical and translational science 11 (3), 305-311, 2018

Simulation Data I – Hypothetical Mathematical Models

Different predictor combinations ($\lambda(\boldsymbol{x})$) for hazard function: $h(t|\mathbf{x}) = h_o(t)e^{\lambda(x)}$

Model	Relationship for covariates in hazard function	Equation of predictors ($\lambda(x)$)
I	Linear	$\beta_1 x_1 + \beta_2 x_2$
II	Nonlinear	$\beta_1 e^{{x_1}^2} + \beta_2 cos{x_2}^2$
III	Interaction	$2x_1 x_2$
IV	Nonlinear + interaction	$\beta_1 e^{{x_1}^2} + \beta_2 cos{x_2}^2 + 2x_1 x_2$
V	Nonlinear + interaction (correlated covariates)	$\beta_1 e^{{x_1}^2} + \beta_2 cos{x_2}^2 + 2x_1 x_2$ $cor(x_1, x_2) = 0.7$
VI	High-dimensional (correlated covariates)	$\sum \beta_i x_i \quad (i = 1,2,\dots,250)$ $cor(x_i, x_j) = 0.7$

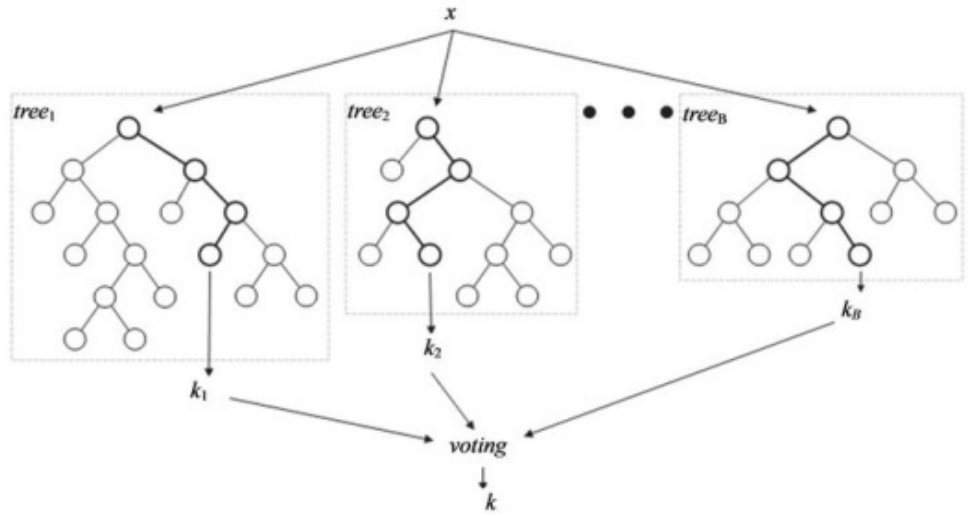
Simulation Data II - Clinically Relevant Models

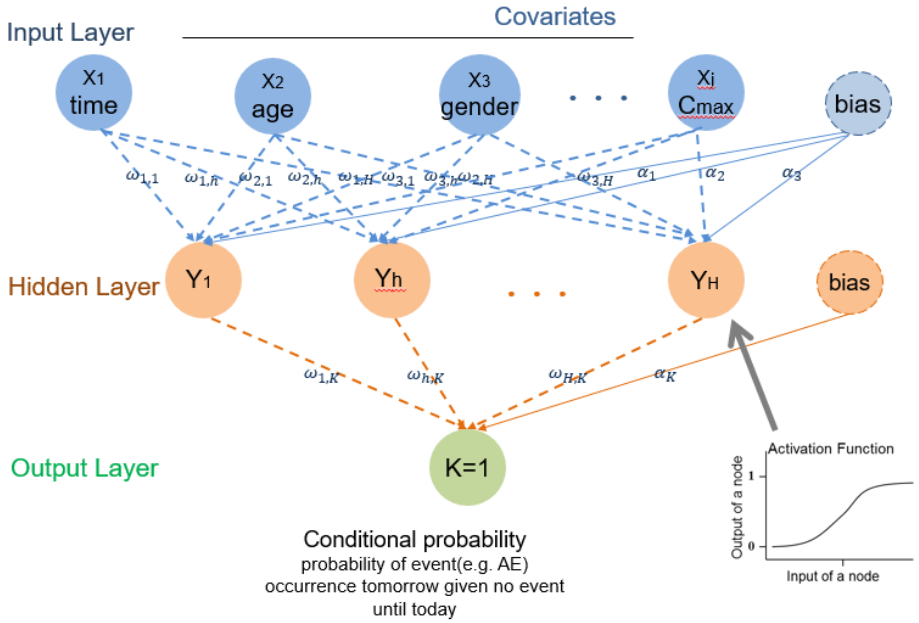
Different predictor combinations ($\lambda(\boldsymbol{x})$) for hazard function: $h(t|\mathbf{x}) = h_o(t)e^{\lambda(x)}$

Model	Relationship for covariates in hazard function	Equation of covariates ($\lambda(\boldsymbol{x})$)
A	Interaction between ECOG and C_{trough}	$\beta_1 \times ECOG + \beta_2 \times Tumor\ size + \beta_3 \times C_{trough} + \beta_{13} \times ECOG \times C_{trough}$
B	Nonlinear drug exposure effects	$\beta_1 \times ECOG + \beta_2 \times Tumor\ size + \beta_{13} \times \frac{60 \times C_{trough}}{30 + C_{trough}}$
C	Interaction between nonlinear drug exposure effect and ECOG	$\beta_1 \times ECOG + \beta_2 \times Tumor\ size + \beta_{13} \times \frac{60 \times C_{trough}}{30 + C_{trough}} \times (1 - ECOG)$

Methods under Evaluation

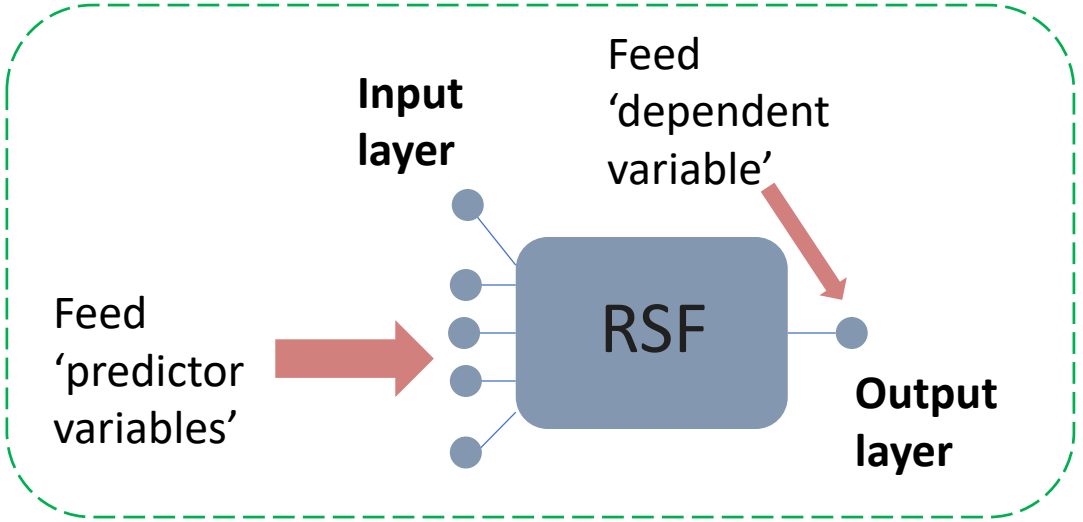
- Random Survival Forest (RSF)
- Artificial Neural Networks (ANN)
- Cox proportional hazard model (Cox)



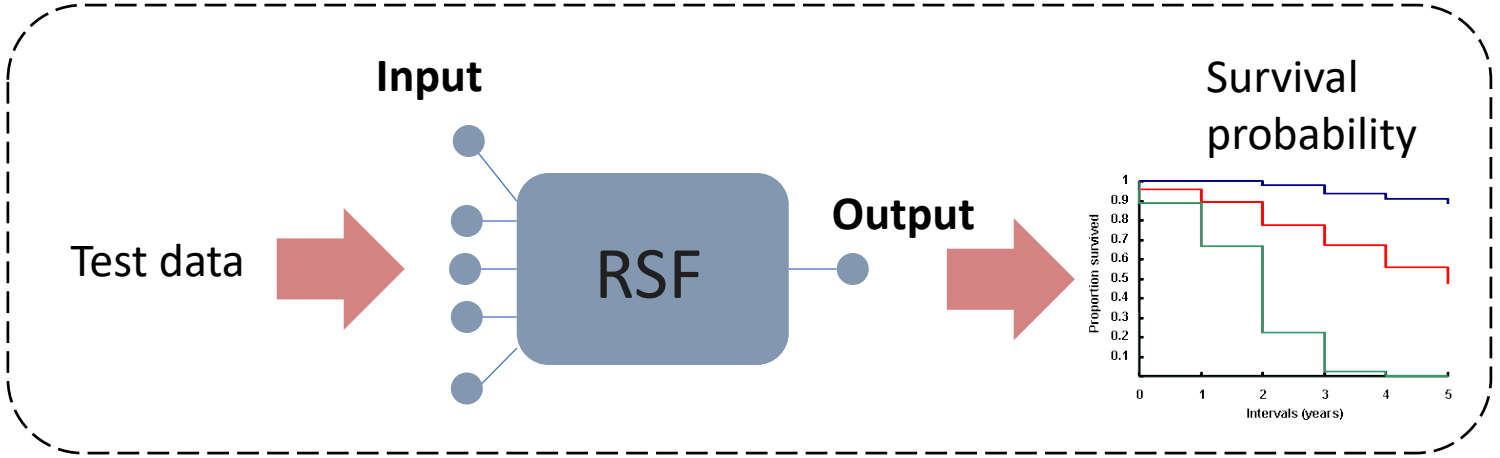


Machine Learning Based Time-to-Event Analysis

Training



Prediction



Predictive Performance Evaluated by the Concordance Index (C-index)

- C-index essentially measures the proportion of '*subject pairs with good predictions*', in which the subject who experiences the event earlier also has the lower predicted survival probability, over all eligible subject pairs.

An example of subject pair with good prediction

Subject	Real Event Time (day)	Predicted Survival Probability
A	10	0.4
B	40	0.9

- C-index = 1; perfect prediction
- C-index = 0.5; random guess

Procedure for performance tests

1. Given a simulation data model (e.g., nonlinear model),
2. One training and one testing data were independently generated,
3. The ML-based methods and the Cox model were applied to the simulated data respectively,
4. The C-index was calculated to assess their predictive performances,
5. Steps 2-4 were repeated 500 times to estimate the mean C-index values.

Results

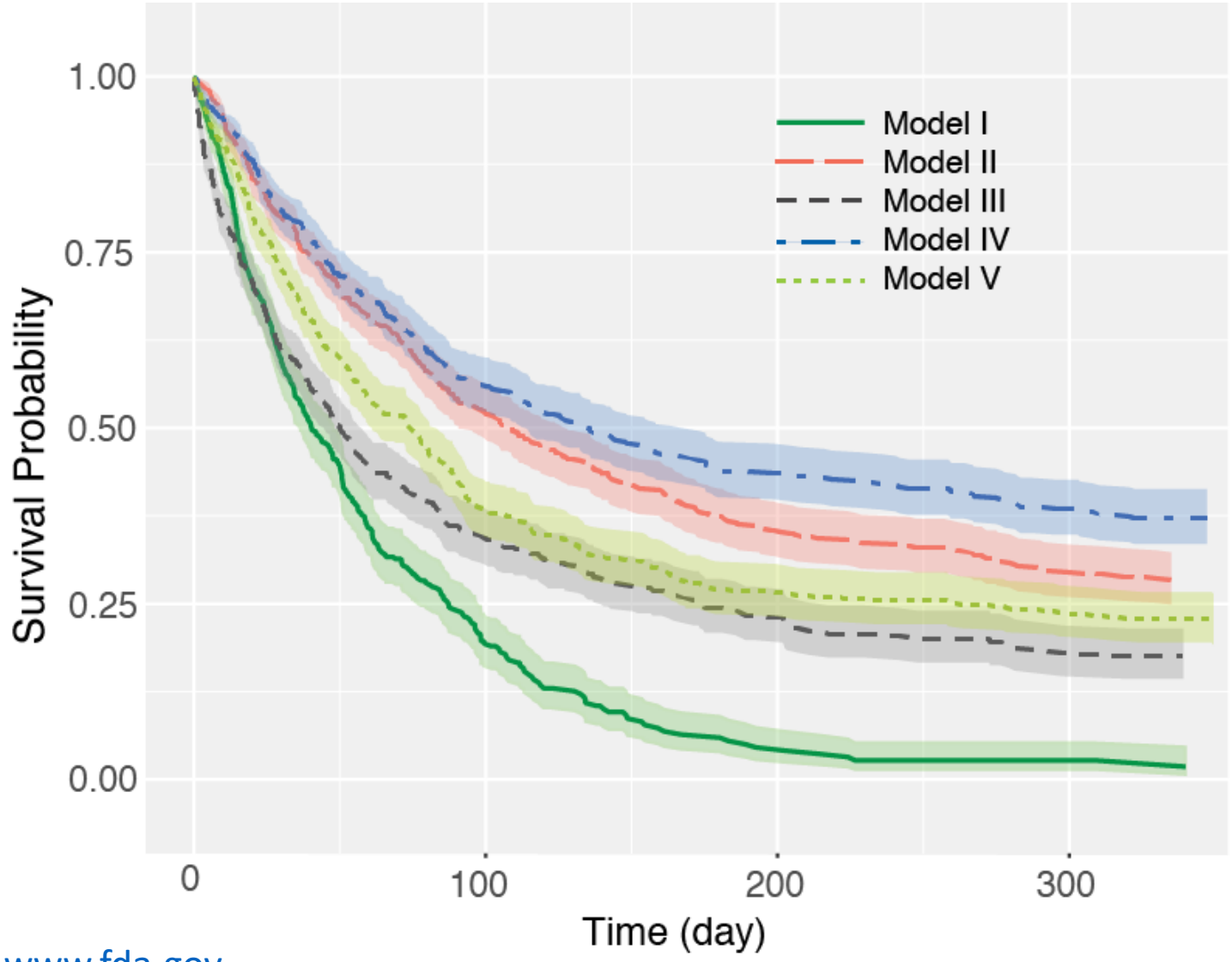
- Simulation Data I
  - Derived from hypothetical mathematical models
- Simulation Data II
  - Derived from clinically relevant models

Simulation Data I – Hypothetical Mathematical Models

Different predictor combinations ($\lambda(\boldsymbol{x})$) for hazard function: $h(t|\mathbf{x}) = h_o(t)e^{\lambda(x)}$

Model	Relationship for covariates in hazard function	Equation of predictors
I	Linear	$\beta_1 x_1 + \beta_2 x_2$
II	Nonlinear	$\beta_1 e^{{x_1}^2} + \beta_2 cos{x_2}^2$
III	Interaction	$2x_1 x_2$
IV	Nonlinear + interaction	$\beta_1 e^{{x_1}^2} + \beta_2 cos{x_2}^2 + 2x_1 x_2$
V	Nonlinear + interaction (correlated covariates)	$\beta_1 e^{{x_1}^2} + \beta_2 cos{x_2}^2 + 2x_1 x_2$ $cor(x_1, x_2) = 0.7$
VI	High-dimensional (correlated covariates)	$\sum \beta_i x_i \quad (i = 1,2,\ldots,250)$ $cor(x_i, x_j) = 0.7$

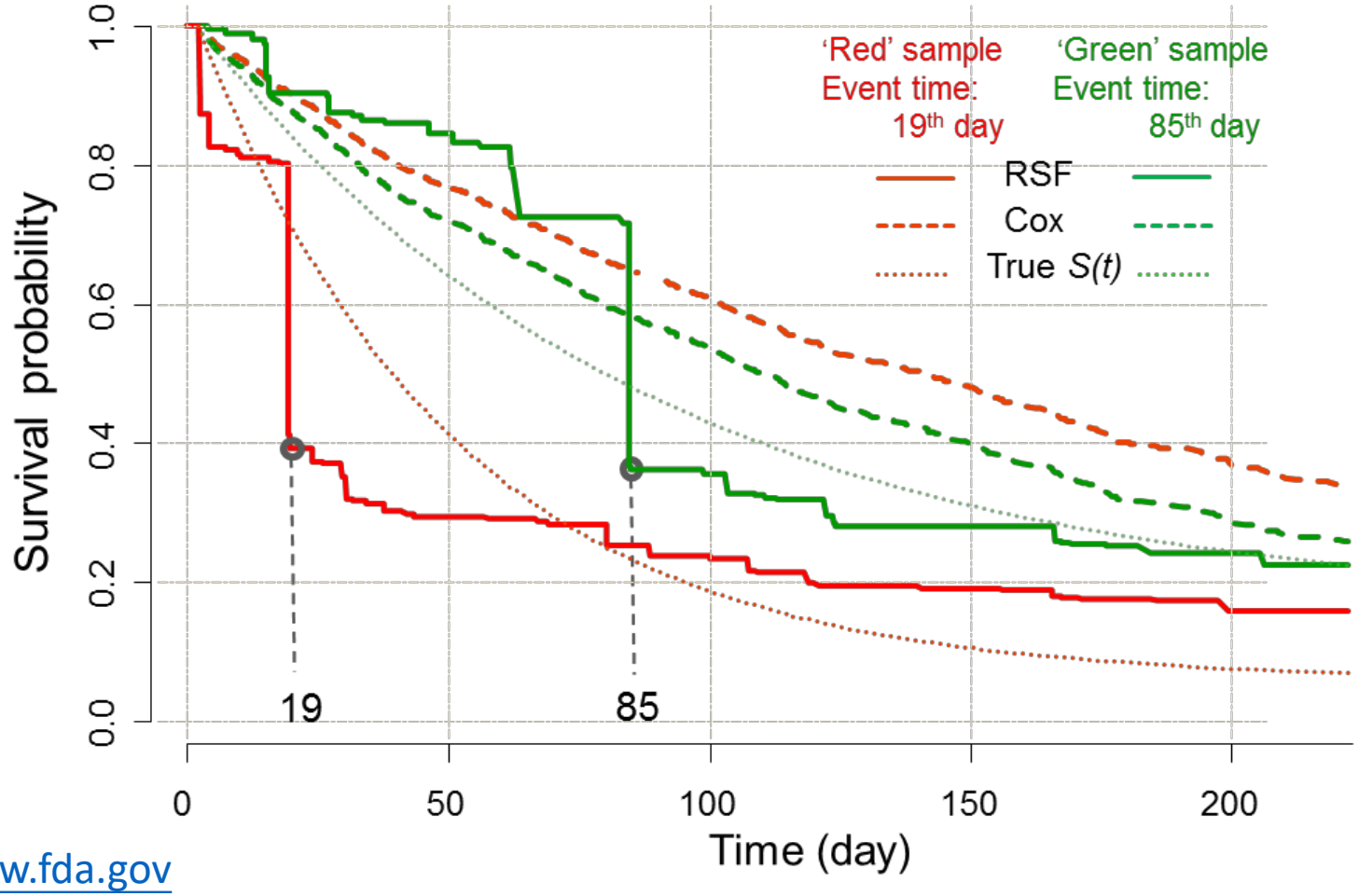
Kaplan-Meier Plots for Simulated Data (Models I-V)



Model	Relationships between predictors
I	Linear
II	Nonlinear
III	Interaction
IV	Nonlinear + interaction
V	Nonlinear + interaction (correlated covariates)

For two exactly same predictors, varying the relationships between them for the hazard function will lead to significantly different survival plots.

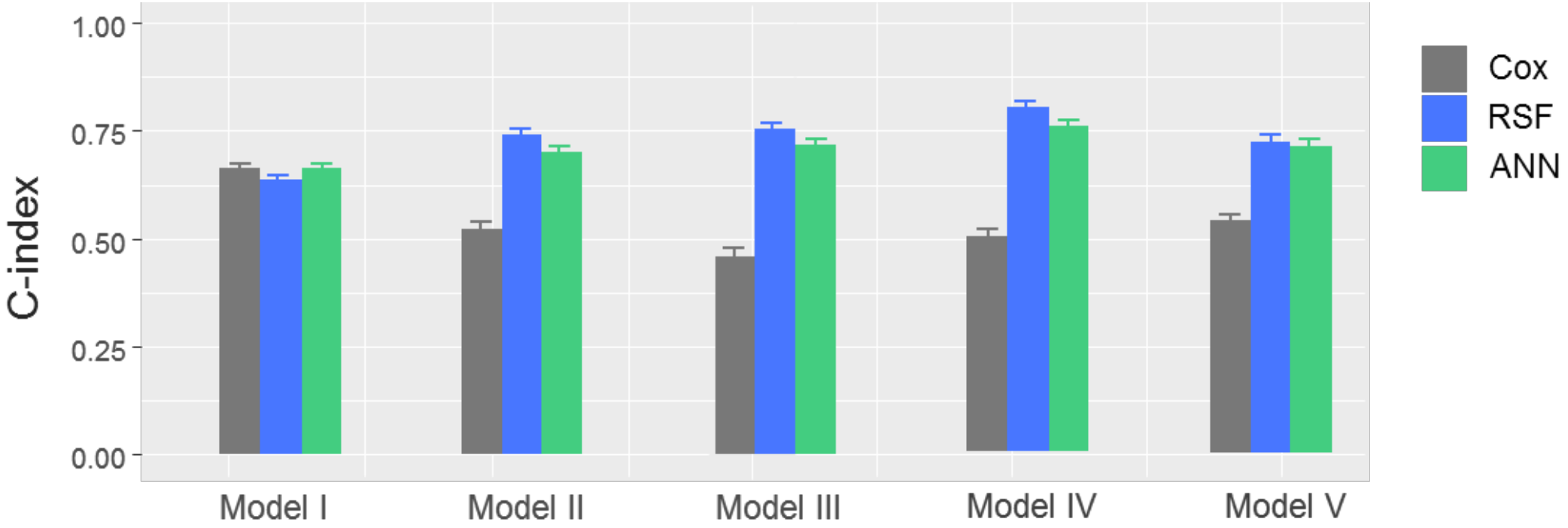
Case example – survival prediction for two hypothetical subjects



Data were generated based on Model II (Nonlinear relationships between predictors)

RSF provides the consistent estimation with the true outcomes, while the Cox model yields the opposite results.

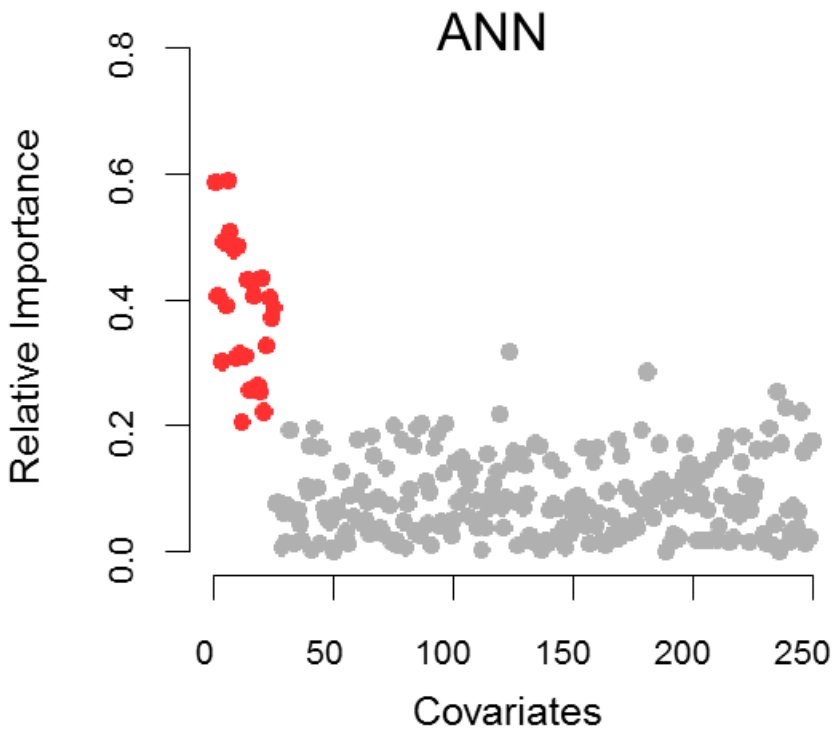
Predictive Performances for Models I-V by C-index

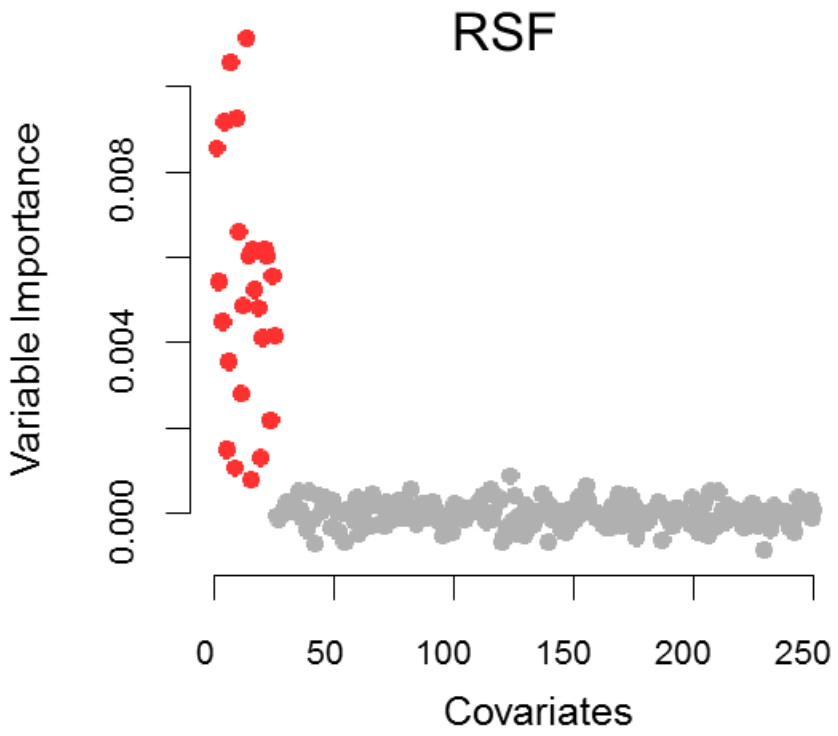


The machine learning based approaches show the significantly better predictive performances than the Cox model, for the data with complex relationships in the hazard functions.

High Dimension Data (Model VI)

- The Cox model failed to yield reasonable estimation due to the parameter identifiability issue.
- The RSF and ANN produced C-index values around 0.71 for predictive performance assessment.
- The ML-based approaches captured influential predictors (i.e., first 25 variables) using the algorithms for variable importance.





Simulation Data II - Clinically Relevant Models

Different covariate combinations ($\lambda(\boldsymbol{x})$) for hazard function: $h(t|\mathbf{x}) = h_o(t)e^{\lambda(x)}$

Model	Relationship for covariates in hazard function	Equation of covariates
A	Interaction between ECOG and C_{trough}	$\beta_1 \times ECOG + \beta_2 \times Tumor\ size + \beta_3 \times C_{trough} + \beta_{13} \times ECOG \times C_{trough}$
B	Nonlinear drug exposure effects	$\beta_1 \times ECOG + \beta_2 \times Tumor\ size + \beta_{13} \times \frac{60 \times C_{trough}}{30 + C_{trough}}$
C	Interaction between nonlinear drug exposure effect and ECOG	$\beta_1 \times ECOG + \beta_2 \times Tumor\ size + \beta_{13} \times \frac{60 \times C_{trough}}{30 + C_{trough}} \times (1 - ECOG)$

Predictive Performances for Models A-C

Predictive Performances by C-index

	Model A	Model B	Model C
Cox	0.50 ± 0.01	0.49 ± 0.02	0.49 ± 0.02
RSF	0.70 ± 0.01	0.59 ± 0.02	0.70 ± 0.01
ANN	0.68 ± 0.03	0.61 ± 0.02	0.69 ± 0.03

- The predictions from the Cox model are like the random guess (C-index ~= 0.5).
- The ML-based approaches (RSF and ANN) offer much more predictive results.

Other applications using machine learning based survival analysis

- *Prediction of the first ANDA submission for NCEs utilizing machine learning methodology*.

https://www.fda.gov/downloads/Drugs/NewsEvents/UCM582150.pdf

- *Revealing association between kinases and adverse events for small-molecule kinase tyrosine inhibitors using machine learning method*.

Presented in the American Society for Clinical Pharmacology and Therapeutics (ASCPT) 2018 annual meeting

Take-Home Message

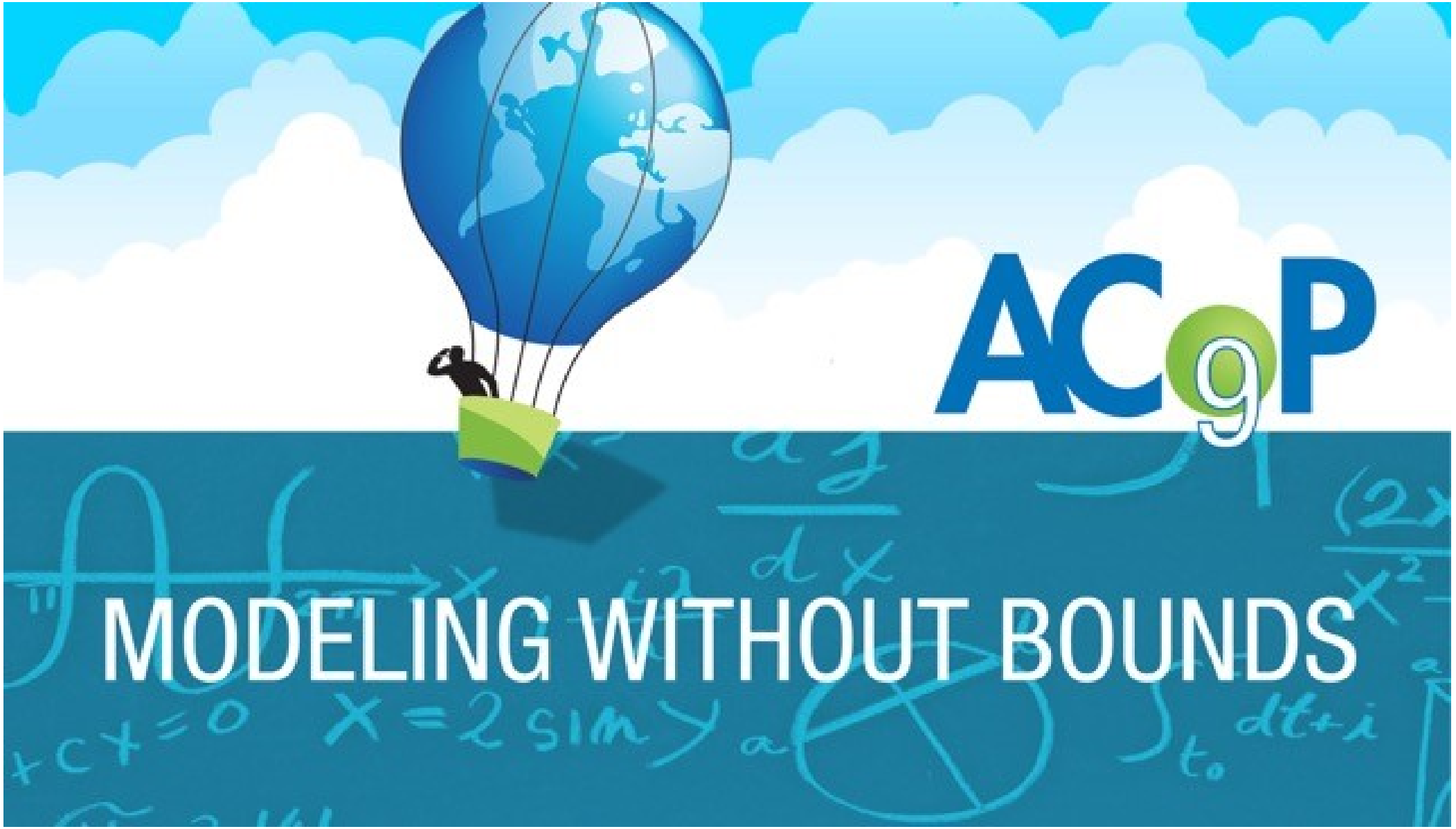
- Our simulation results show that the machine learning based survival analysis outperform the conventional Cox model for complex survival data:
  - Nonlinear
  - High-order
  - High-dimension

- Our study suggests that the machine learning based analysis can be the useful complement to the Cox model for time-to-event analysis when data show the high degree of complexity.

Acknowledgement

- Liang Zhao
- Xiajing Gong
- Robert Lionberger
- Andrew Babiskin
- FDA Critical Path Initiative (https://www.fda.gov/scienceresearch/specialtopics/criticalpathinitiative/default.htm)

www.fda.gov



San Diego, California

October 7 – 11, 2018