

# Deep Kernel Survival Analysis and Subject-Specific Survival Time Prediction Intervals

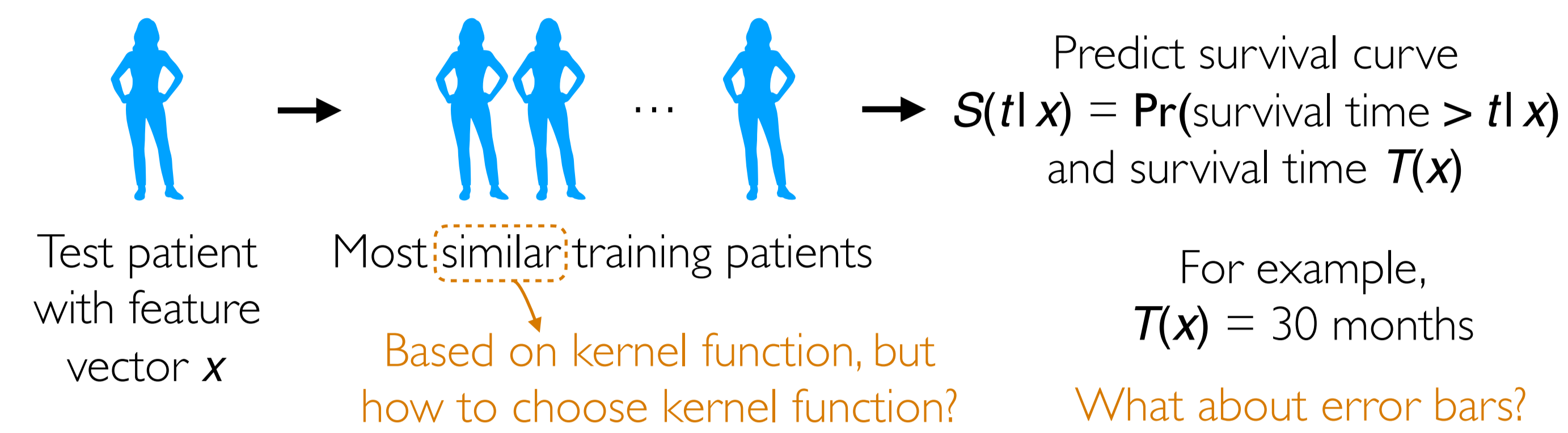
George H. Chen

Carnegie Mellon University  
**Heinz College**  
INFORMATION SYSTEMS • PUBLIC POLICY • MANAGEMENT

## Introduction

### Kernel survival analysis methods

Predict time-to-event outcome (e.g., time until death) using a *kernel function* (measures how similar any two patients are)



**This paper's contributions:**

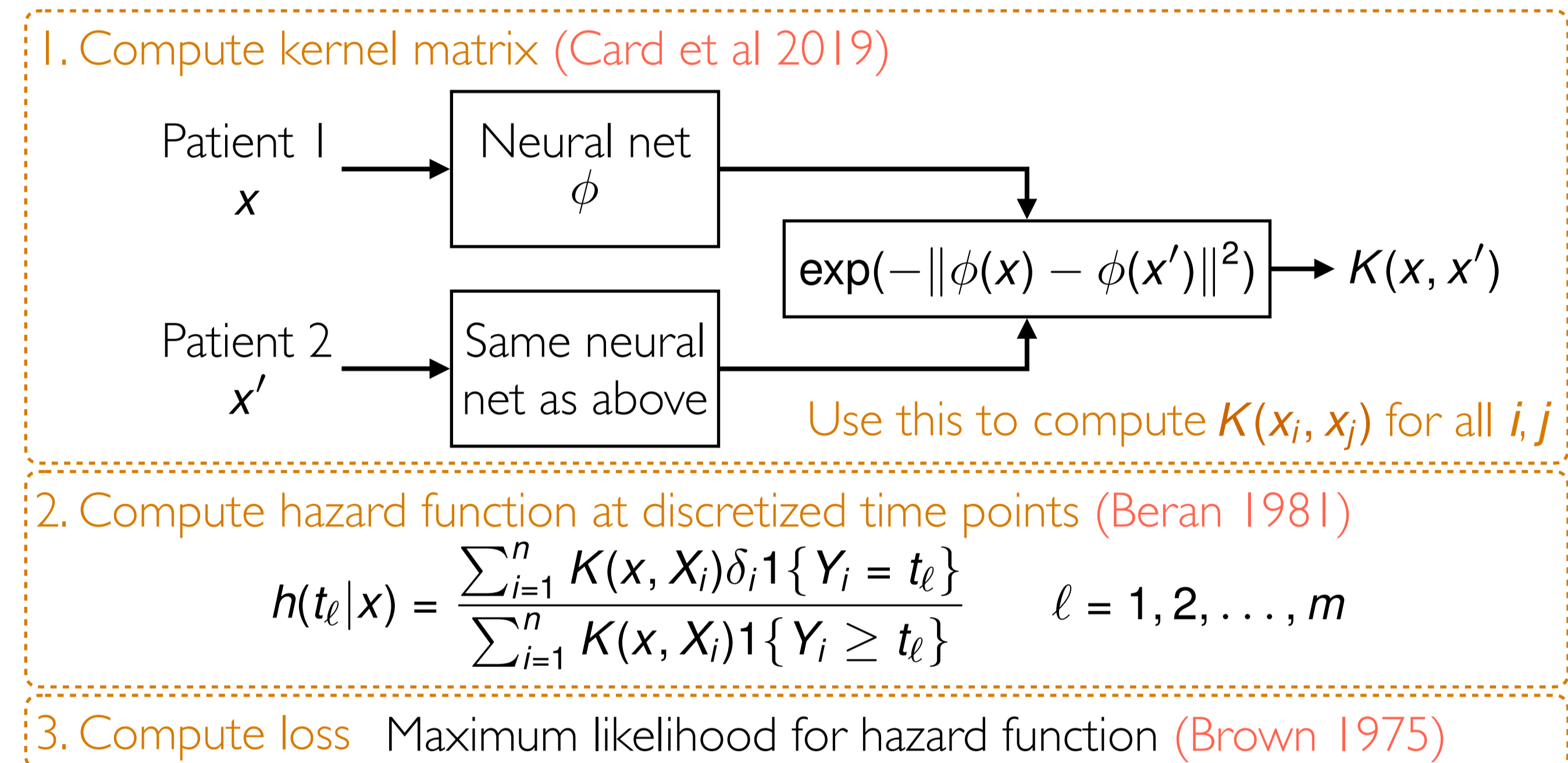
- (1) Automatic kernel function learning for survival analysis using a neural net framework
- (2) Construct statistically valid subject-specific survival time prediction intervals

## Deep Kernel Survival Analysis

Training data:  $(X_1, Y_1, \delta_1), \dots, (X_n, Y_n, \delta_n)$  Standard survival analysis setup

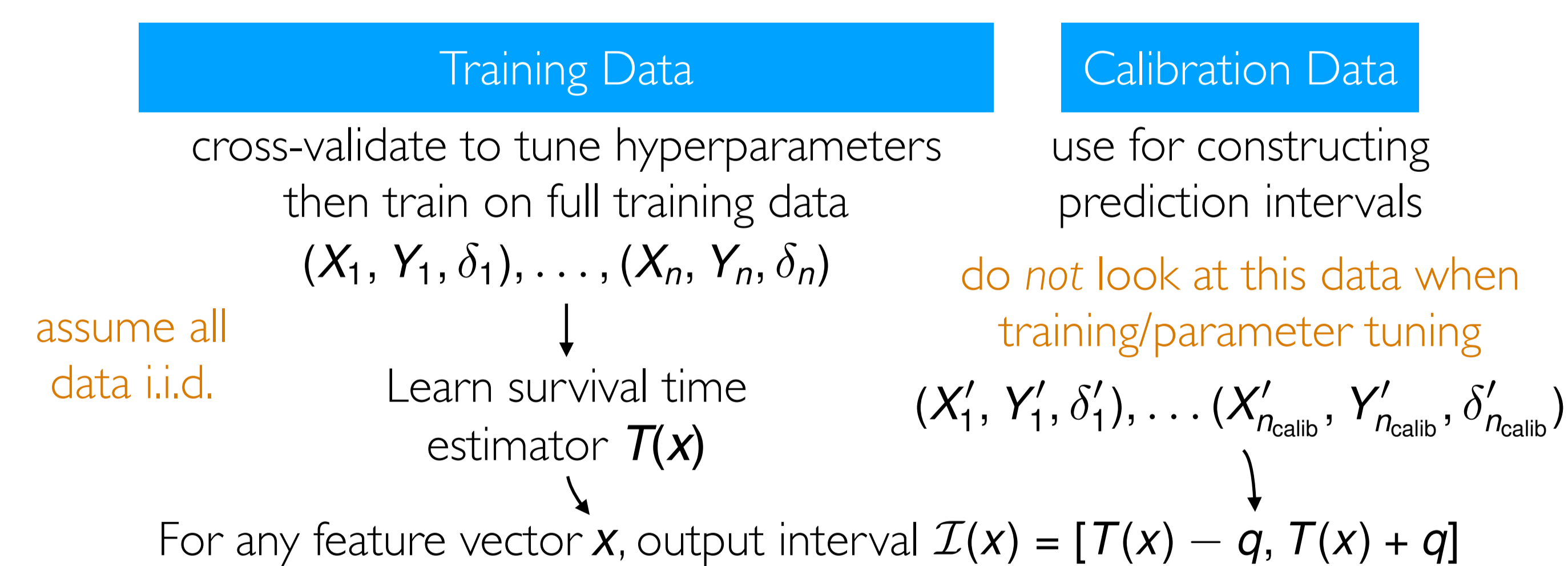
Discretize time:  $t_1 < t_2 < \dots < t_m$  For background, see my CHIL 2020 tutorial: <https://sites.google.com/view/chil-survival>

### High-level strategy:



## Survival Time Prediction Intervals

Use **split conformal prediction** and its kernel variant (Papadopoulos et al 2002, Lei et al 2015, Tibshirani et al 2019)



**Marginal coverage:** sample test feature vector  $X$  the same way as training data; 95% of the time, true survival time  $X$  is in  $\mathcal{I}(X)$

Like confidence intervals but for prediction instead of inference!

For simplicity I am using 95%; more generally can use  $1 - \alpha \in (0, 1)$

Higher target coverage level  $\Rightarrow$  wider prediction intervals

How to compute the error bar  $q$  to satisfy marginal coverage:

1. Compute "residuals" for calibration data:
 
$$R_i = \begin{cases} |Y'_i - T(X'_i)| & \text{if } \delta'_i = 1 \\ \max\{Y'_i - T(X'_i), 0\} & \text{if } \delta'_i = 0 \end{cases} \quad \text{for } i = 1, 2, \dots, n_{\text{calib}}$$

predict death before censoring time

Insert a last residual of infinity:  $R_{m+1} = \infty$
2. The residuals  $R_1, R_2, \dots, R_{m+1}$  form an empirical distribution; let  $q$  be the 95% percentile of this distribution (could be  $\infty$ )

Prediction interval of test subject's survival time *relative to how similar test subject is to another individual* is also possible using kernel function (local coverage)

## Numerical Experiments

**Datasets:** standard healthcare time-until-death datasets

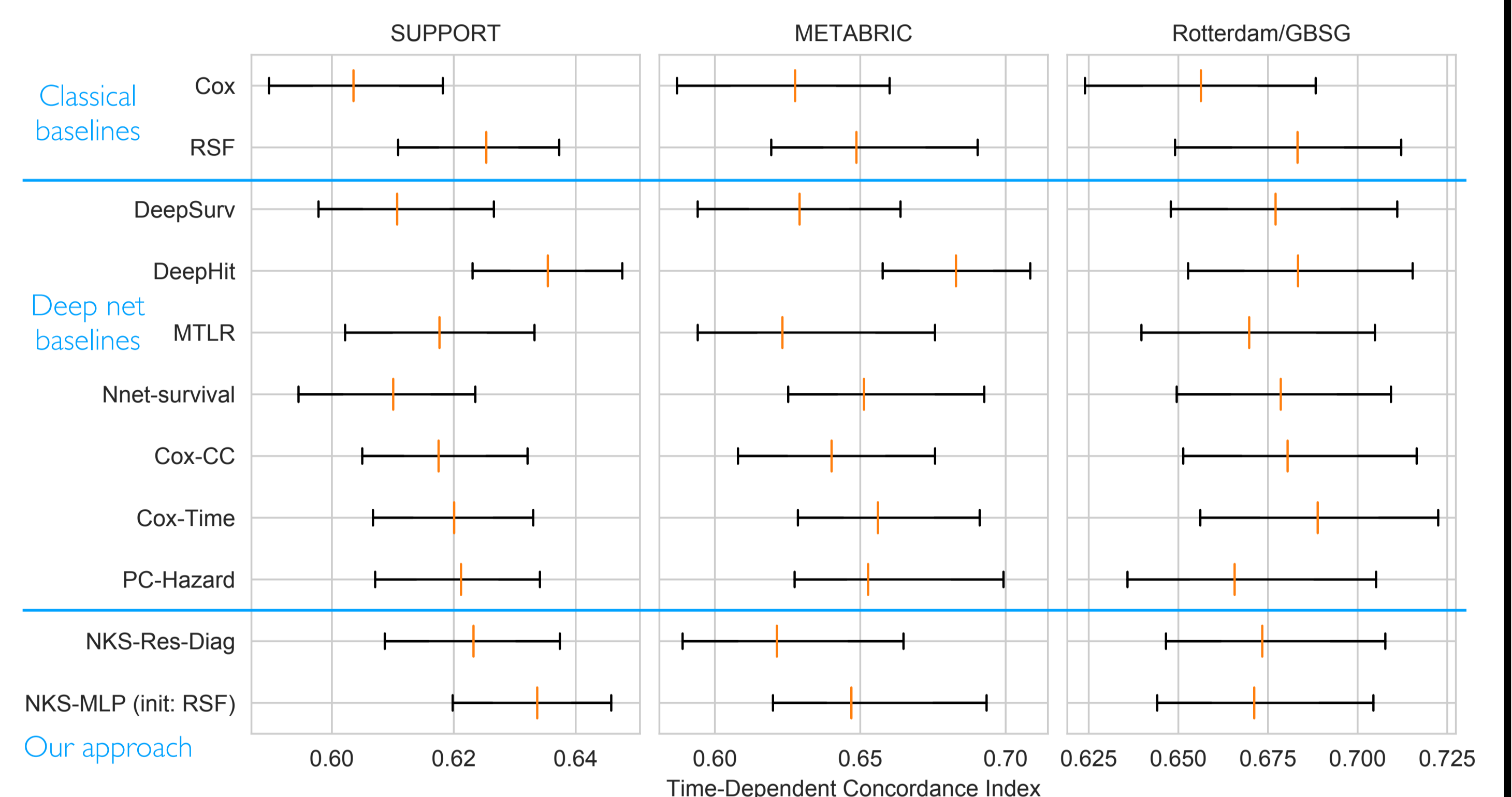
Dataset	# subjects	# features	% censored	Observed times (min / median / max)
SUPPORT (Knaus et al 1995)	8873	14	32.0%	0.10 / 7.59 / 66.70 months
METABRIC (Curtis et al 2012)	1904	9	42.1%	0 / 114.90 / 355.20 months
Rotterdam (Foekens et al 2000)	1546	7	37.4%	1.25 / 44.75 / 84 months
GBSG (Schumacher et al 1994)	686	7	56.4%	0.26 / 35.61 / 87.36 months

SUPPORT, METABRIC: for each, do randomized 70/30 train/test split  
For Rotterdam & GBSG: train on Rotterdam, test on GBSG

### Experiment: Benchmark using concordance index

For every algorithm (including ours): 5-fold CV to tune hyperparameters, train on full training data, evaluate on test data

Error metric: time-dependent concordance index (Antolini et al 2015), like AUC for survival data (values 0 to 1, with 1 being best)—ranking accuracy



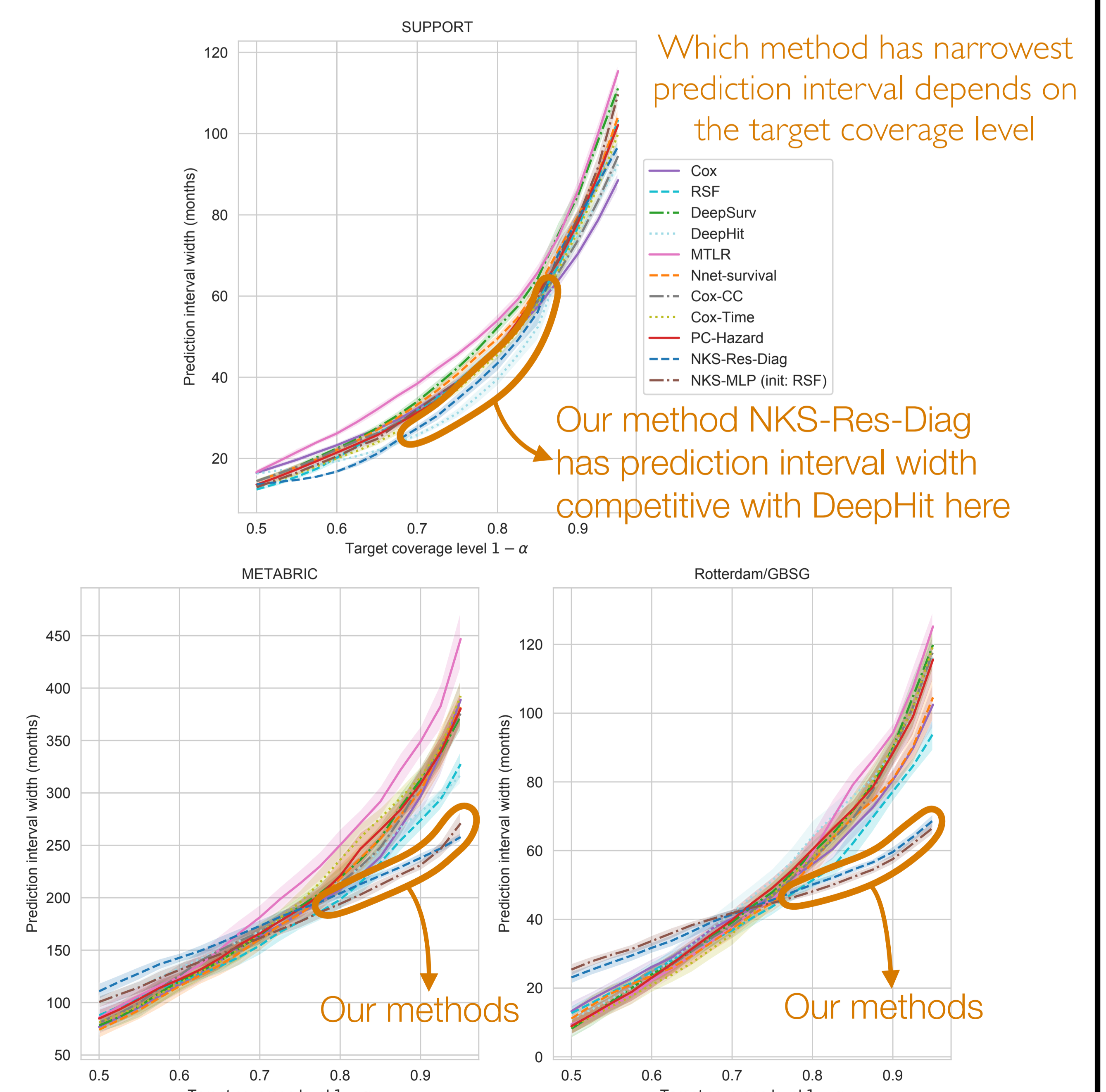
Line segments show 95% bootstrap confidence intervals

Deep kernel survival analysis is able to achieve competitive accuracy with deep learning baselines

### Experiment: Benchmark using prediction interval widths

Marginal prediction interval width: gauges both accuracy and uncertainty

Variant of above experiment: split test data into 50/50 calibration/actual test so that we can construct survival time prediction intervals using calibration data



Prediction interval widths are a new performance metric for survival methods!  
More experiments can be found in the paper