Removing Race from the Kidney Donor Risk Index

Aleisha Khan
MS, Physiology, Georgetown University
General Assembly Data Science Immersive EC 13

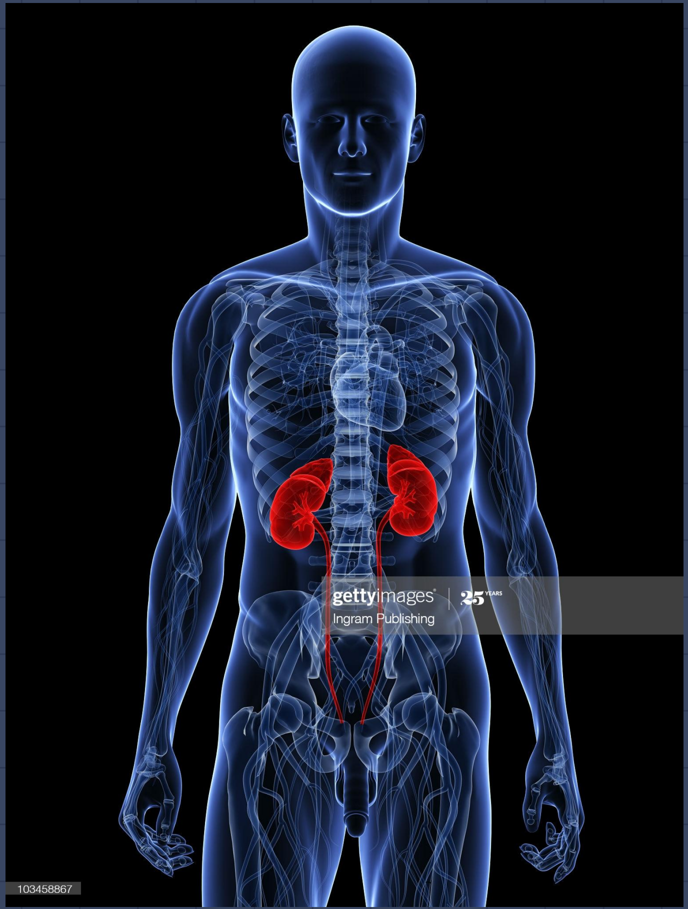
Kidney Transplants in the U.S.

Kidney Failure - when the kidneys lose the ability to sufficiently filter waste/toxins from your blood

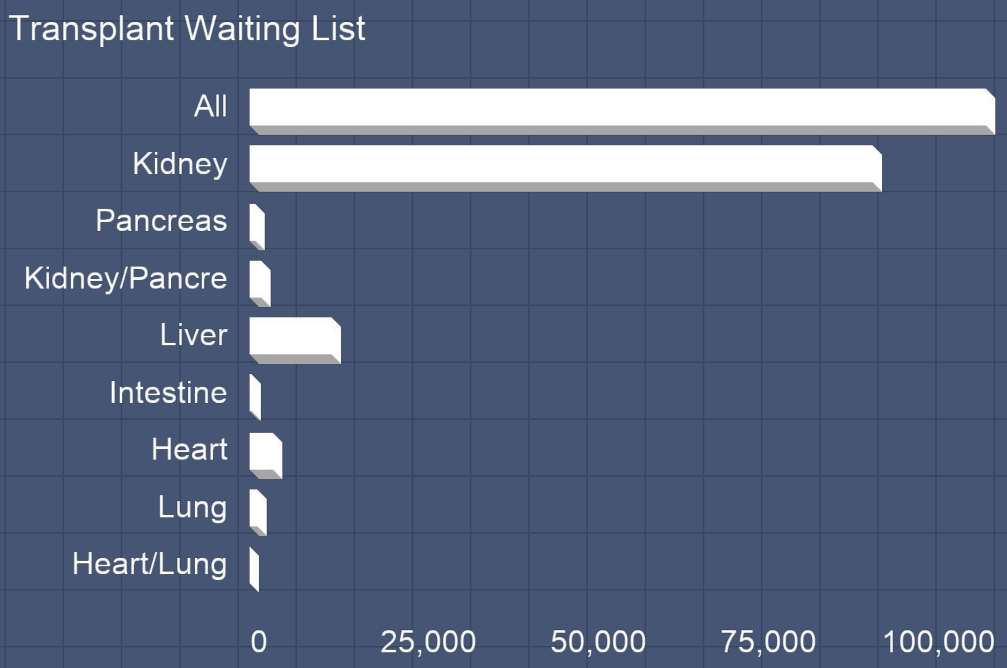
The two leading causes of kidney failure in the United States are:

1.) Diabetes (34.2 million Americans)

2.) High Blood Pressure (108 million Americans)

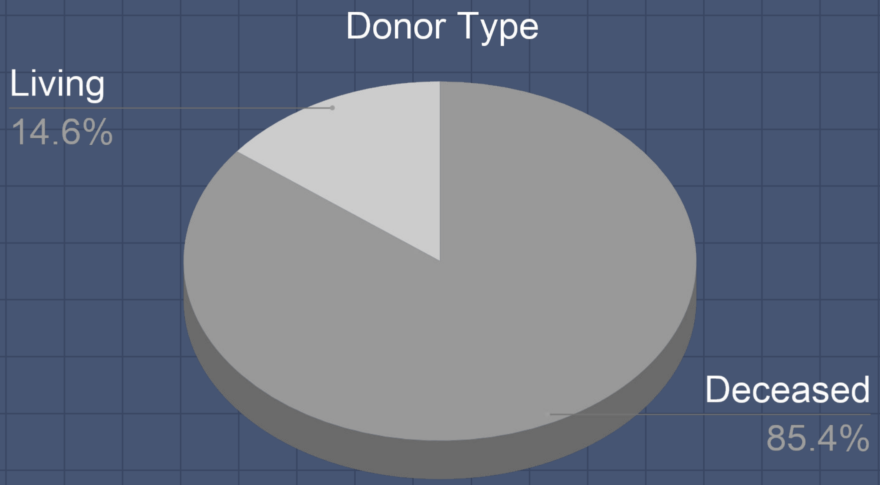


Kidney Transplants in the U.S.



Waiting list candidates as of Dec 10, at 8:53 pm (via OPTN)

Kidney Transplants in the U.S.



Transplants performed January-October 2020

A Crash Course in Kidney Transplant Terminology

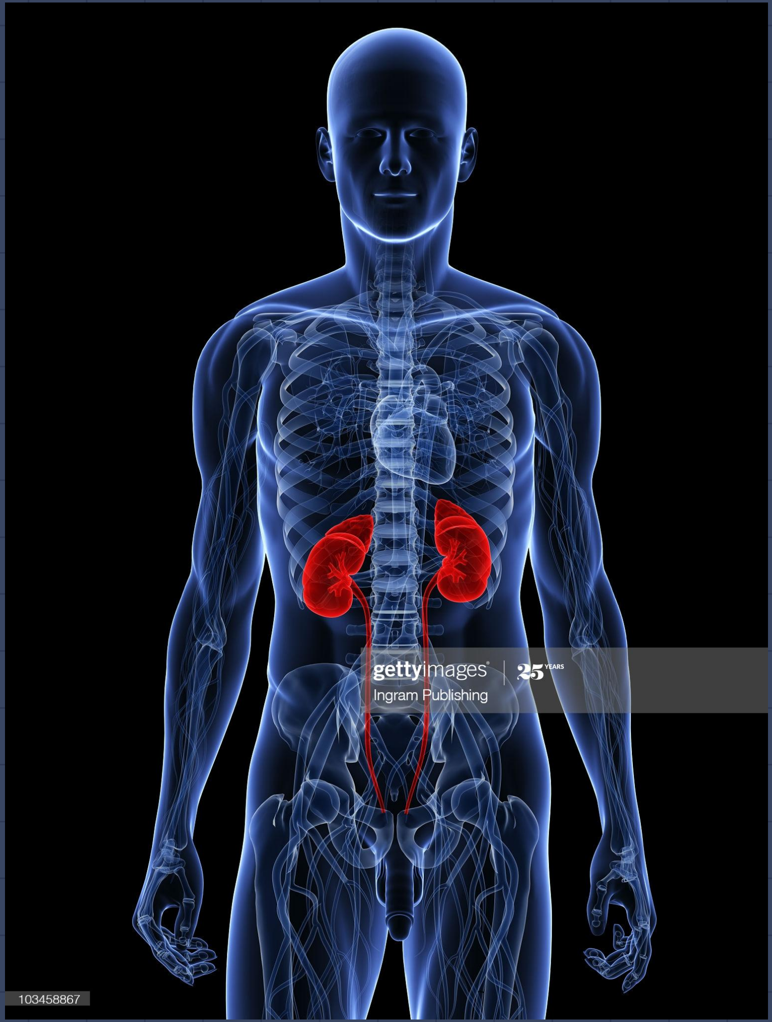
Deceased Donor - an organ donor who has undergone brain or circulatory death prior to donation

ECD - expanded criteria donor

Graft failure - when the transplanted organ does not function correctly, resulting in the need for dialysis or a new transplant

Can be caused by:

- Blood clot
- Fluid collection
- Infection
- Acute or Chronic Rejection



What is the Kidney Donor Risk Index?

ORGAN PROCUREMENT AND TRANSPLANTATION NETWORK:

"The Kidney Donor Risk Index (KDRI) combines a variety of donor factors to summarize the risk of graft failure after kidney transplant into a single number"

KDRI Pros and Cons

Pros

- Reduced waiting list time for critically ill patients willing to accept a high KDPI kidney
- Greater granularity than the previously used ECD/SCD designation
- Useful to determine whether to accept an offer of both kidneys versus if only 1 kidney is available

Cons

- Based on data from kidney transplants between 1995- 2005
- Original algorithm defined graft failure to include death, regardless of cause
- Race adjustment returns a higher risk of graft failure for kidneys from black kidney donors

"…despite mounting evidence that race is not a reliable proxy for genetic difference, the belief that it is has become embedded, sometimes insidiously, within medical practice. One subtle insertion of race into medicine involves diagnostic algorithms and practice guidelines that adjust or "correct" their outputs on the basis of a patient's race or ethnicity. Physicians use these algorithms to individualize risk assessment and guide clinical decisions. By embedding race into the basic data and decisions of health care, these algorithms propagate race-based medicine. Many of these race-adjusted algorithms guide decisions in ways that may direct more attention or resources to white patients than to members of racial and ethnic minorities."

Vyas, Darshali A., Leo G. Eisenstein, and David S. Jones. "Hidden in plain sight—reconsidering the use of race correction in clinical algorithms." (2020): 874-882.

THE DATA

Organ Procurement and Transplantation Network (OPTN)
Standard Transplant Analysis and Research (STAR)

1

The Data

Types of Data

- Donor and Recipient Biochemical Markers
- Demographic Data
- Follow up data (extensive)
- Location

Scope of Data

- Originally in SAS format
- >18 Gb uncompressed
- Relatively high data integrity

Exclusion Criteria

- <18y.o.
- Multi-organ transplant
- Previous transplant
- ABO incompatible
- Living Donor

STAR Files

	Waitlist	Deceased Donor	Follow-Up
Number of records	521,064	233,354	3,813,134
Number of features	474	500	62

SQLite

Used to extract data from original, full STAR database; implemented exclusion and inclusion criteria, exported for use in Python.

Python

Used to extract, transform, and load data from SQLite; feature selection, statistical analyses to decrease feature space (1714 -> 100), model building

Google Cloud AI platform

Used for custom model training, Bayesian hyperparameter optimization

Final Dataset (after exclusion)

Number of Transplant Records: 102,480

Number of Features: 100

Date Range: 2005-2020

CHOOSING THE MODEL

Random Survival Forests

2

Survival Modeling
(or time-to-event analysis)

For censored data (dropout, death, failure, 'event')

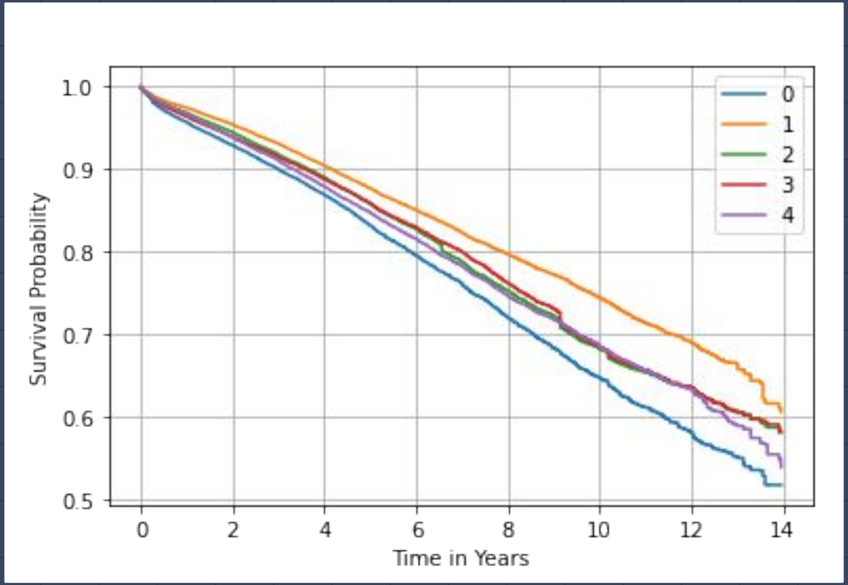
= Graft failure

Allows use of all records even if lost to follow up/event never occurred

Follows patients post-transplant to see when graft fails or lost to follow up

Prediction = probability of event occuring within time period

Useful for predicting hazard ratio for health/demographic/location factors



Why Random Survival Forests?

Cox Regression (Original KDRI)

Estimates log-risk function as a linear combination of static covariates and baseline hazard

- Proportional Hazards Constraint:

  The hazard functions for any two subjects stay proportional at any point in time and the hazard ratio does not vary with time.

Random Survival Forests

Constructs ensemble estimate for the cumulative hazard function from base learners

- Can account for change in risk over time
- Minimizes Overfitting

TRAINING THE MODEL

Using Google Cloud AI Platform

3

Model Parameterization

Training Job

Created custom training job on GCP

`scikit-survival`

Variables to Tune

`n_estimators`

`min_samples_split`

`min_samples_leaf`

`max_features`

Accuracy Metric

Concordance Index

Where the rubber hits the road

	Trial 1	Trial 2	Trial 3
Time To Run	39 hours	10 hours	23 hours
Number of Features	100	100	33
Best Concordance	0.634	0.621	0.634
Cost	~$140	~$13	~$22

Cloud Parameters

HyperTune trials

Trial ID	concordance_index ↓	Training step	Elapsed time	n_estimators	min_samples_split	min_samples_
12	0.62131	1,000	2 hr 46 min	300	0.05	0.05
7	0.62089	1,000	2 hr 3 min	200	0.05	0.05
9	0.62077	1,000	2 hr 59 min	300	0.1	0.05
2	0.61873	1,000	59 min 39 sec	100	0.05	0.05
8	0.61852	1,000	51 min 31 sec	100	0.1	0.05
5	0.61832	1,000	1 hr 53 min	200	0.1	0.05
6	0.60931	1,000	34 min 16 sec	100	0.05	0.1
1	0.60807	1,000	1 hr 7 min	200	0.1	0.1
4	0.60725	1,000	1 hr 4 min	200	0.05	0.1
10	0.60664	1,000	1 hr 35 min	300	0.05	0.1

Rows per page: 10 1 – 10 of 12

`n_estimators = {100, 200, 300}` `min_samples_split = {0.05, 0.1}` `min_samples_leaf = {0.05, 0.1}`

MODEL INTERPRETATION AND RECURSIVE FEATURE SELECTION

Using the eli5 package in Python

4

`eli5`: "Explain Like I'm 5"

Used to extract feature importance

- 10% of data, 5-fold cross-validation (~10,000 rows)

Predicts feature importance by measuring change in performance metric (concordance index) after shuffling one column at a time

Outcome

Weight	Feature
0.0195 ± 0.0135	AGE_DON
0.0118 ± 0.0152	AGE
0.0092 ± 0.0243	INIT_AGE
0.0071 ± 0.0074	Transplant_Year
0.0060 ± 0.0093	ECD_DONOR
0.0020 ± 0.0047	HIST_HYPERTENS_DON
0.0012 ± 0.0011	BUN_DON
0.0011 ± 0.0009	URINE_INF_DON
0.0009 ± 0.0023	DRMIS_2.0
0.0008 ± 0.0035	COLD_ISCH_KI
0.0007 ± 0.0009	HIST_CIG_DON
0.0005 ± 0.0016	DISTANCE
0.0005 ± 0.0023	HTLV_DON_ND
0.0003 ± 0.0009	REGION_2
0.0003 ± 0.0007	PO2_DON
0.0002 ± 0.0031	CREAT_TRR
0.0001 ± 0.0002	RDR1_4.0
0.0001 ± 0.0005	DIABETES_DON
0.0001 ± 0.0010	CREAT_DON
0.0001 ± 0.0010	INOTROP_SUPPORT_DON
0.0001 ± 0.0002	EBV_SEROSTATUS_P
0.0001 ± 0.0004	DEATH_MECH_DON_3
0.0001 ± 0.0003	DR1_4
0.0001 ± 0.0013	BMI_CALC
0.0001 ± 0.0005	HEMATOCRIT_DON
0.0001 ± 0.0009	HIST_OTH_DRUG_DON
0.0000 ± 0.0021	PH_DON
0.0000 ± 0.0003	EDUCATION_5.0
0.0000 ± 0.0001	DIAG_KI_3040.0
0.0000 ± 0.0004	DEATH_CIRCUM_DON_2
0.0000 ± 0.0000	REGION_5
0.0000 ± 0.0001	RDR1_13.0
0.0000 ± 0.0000	DR1_13
-0.0000 ± 0.0001	NON_HRT_DON
-0.0000 ± 0.0004	SODIUM170_VAL_DON
-0.0000 ± 0.0001	EBNA_DON_P
-0.0000 ± 0.0002	DEATH_CIRCUM_DON_997
-0.0000 ± 0.0001	FUNC_STAT_TRR_2090.0
-0.0000 ± 0.0001	HIST_COCAINE_DON
-0.0000 ± 0.0035	INIT_WGT_KG
-0.0001 ± 0.0005	PO2_FIO2_DON
-0.0001 ± 0.0012	COD_CAD_DON_1
-0.0001 ± 0.0004	FUNC_STAT_TRR_2070.0
-0.0001 ± 0.0007	TATTOOS_DON
-0.0001 ± 0.0007	TATTOOS
-0.0001 ± 0.0003	DEATH_CIRCUM_DON_5
-0.0001 ± 0.0004	PCO2_DON
-0.0001 ± 0.0003	CDC_RISK_HIV_DON
-0.0001 ± 0.0005	WGT_KG_DON_CALC
-0.0001 ± 0.0026	HTLV_DON_N
-0.0001 ± 0.0003	FUNC_STAT_TCR_2070.0
-0.0001 ± 0.0003	CLIN_INFECT_DON
-0.0002 ± 0.0033	WGT_KG_CALC
-0.0002 ± 0.0018	SGOT_DON
-0.0002 ± 0.0008	HLAMIS_6.0
-0.0003 ± 0.0048	COD_CAD_DON_2
-0.0003 ± 0.0018	DAYSWAIT_ALLOC
-0.0004 ± 0.0053	DEATH_MECH_DON_11
-0.0005 ± 0.0025	INIT_HGT_CM
-0.0006 ± 0.0023	WGT_KG_TCR
-0.0006 ± 0.0020	SGPT_DON
-0.0006 ± 0.0020	HGT_CM_TCR
-0.0009 ± 0.0013	HGT_CM_CALC

Final Results for Feature Selection

1. Donor Age
2. Recipient Age
3. Age Waitlisted for Transplant
4. Transplant Year
5. ECD
6. History of Hypertension (Donor)
7. BUN (Donor)
8. Urine infection (Donor)
9. D Locus Mismatch
10. Cold Ischemic Time
11. History of Cigarette Use (Donor)
12. Distance (kidney → transplant center)
13. Human T-Lymphotrophic Virus (Donor)
14. UNOS Region
15. PO2 (Donor)
16. Creatinine (Recipient)
17. Recipient-DR1 Antigen
18. History of Diabetes (Donor)
19. Inotropic Support (Donor)
20. EBV Serostatus(Donor)
21. Mechanism of Death (Donor)
22. Candidate DR1 Antigen from Waiting List
23. BMI (Recipient)
24. Hematocrit (Donor)
25. History of Other Drug Use (Donor)

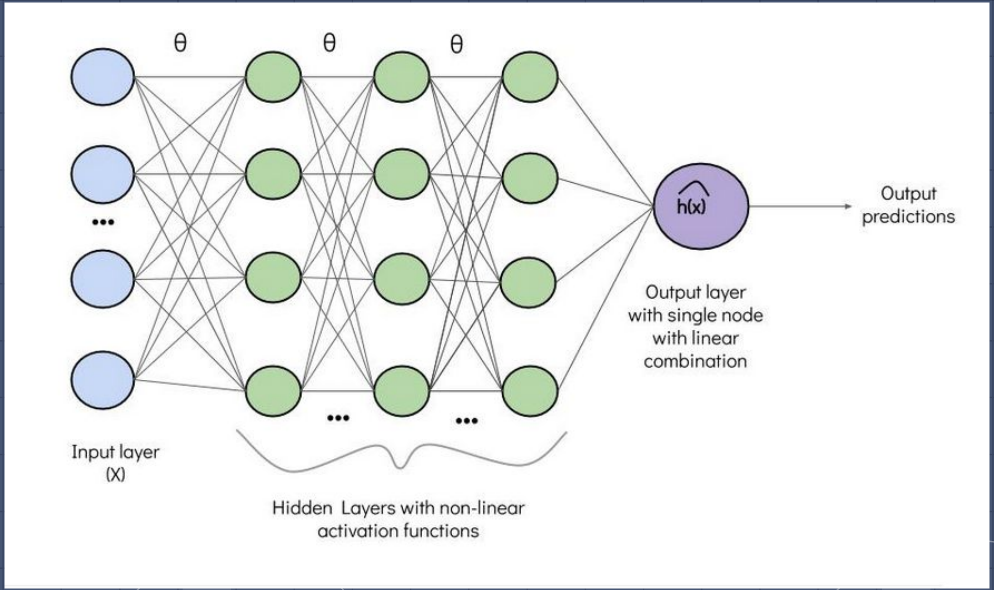
DeepSurv

DeepSurv

Deep Learning, feed-forward neural net for survival analysis

Pycox

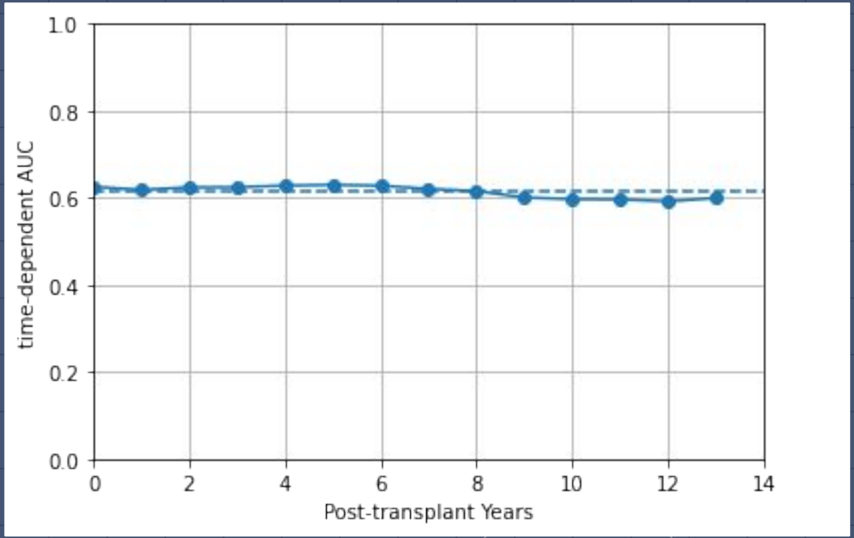
Parameters: 256 nodes x 4 layers, Dropout = 0.4, Optimizer = Adam, batch_size = 128



DeepSurv vs Random Survival Forests

	RSF	DeepSurv	Original Model
Concordance Index	0.634	0.634	0.62
Integrated Brier Score (5 years)	0.05	0.06	N/A
Integrated Brier Score(10 years)	0.12	0.15	N/A

Cumulative dynamic AUC for RSF



Recommended Resources

- Deep Learning for Survival Analysis, Humboldt-Universitat Zu Berlin: https://humboldt-wi.github.io/blog/research/information_systems_1920/group2_survivalanalysis/#introduction_sa
- Hyperparameter Tuning On Google Cloud Platform With Scikit-Learn: https://towardsdatascience.com/hyperparameter-tuning-on-google-cloud-platform-with-scikit-learn-7d6155195efb
- Scikit-Survival Documentation: https://scikit-survival.readthedocs.io/en/latest/
- Pycox Documentation: https://github.com/havakv/pycox

Thank you so much!!!