



UiT The Arctic University of Norway

SURVIVAL ANALYSIS

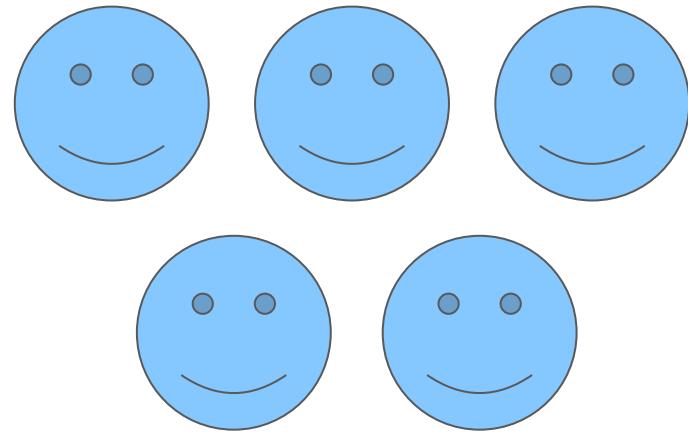
An introduction to basic concepts

Rafael Nozal Cañadas

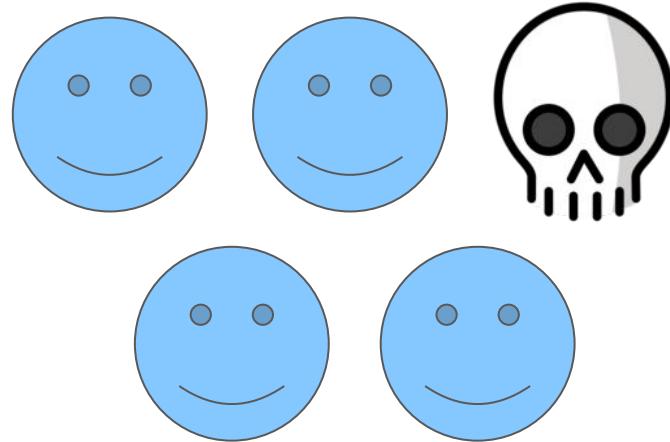
HDL Seminar 2024.01.29



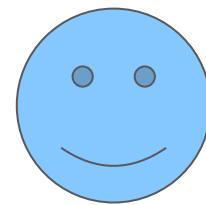
1 - THE BASICS



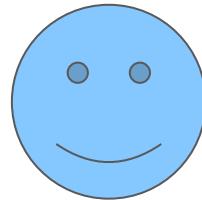
3 years

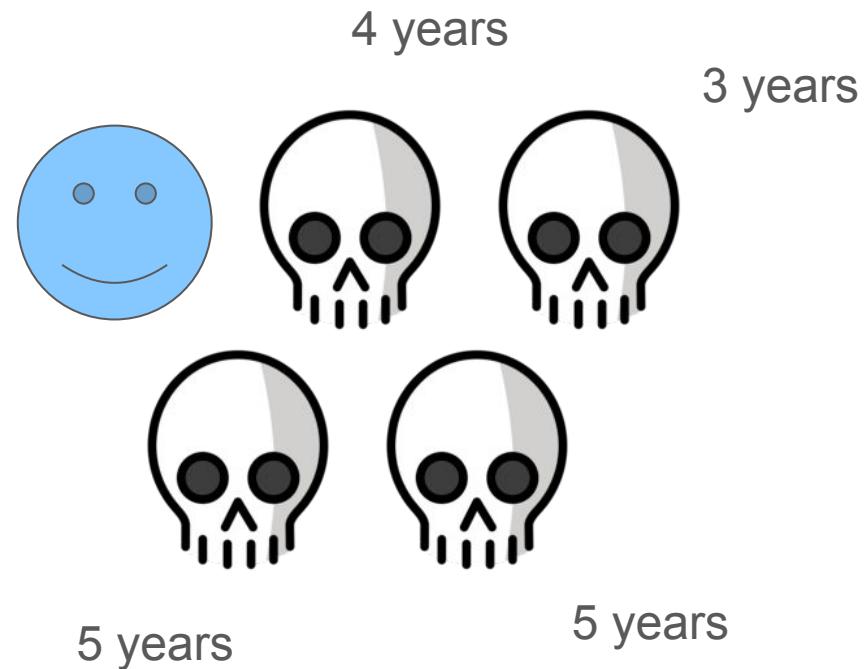


4 years



3 years





8 years

4 years

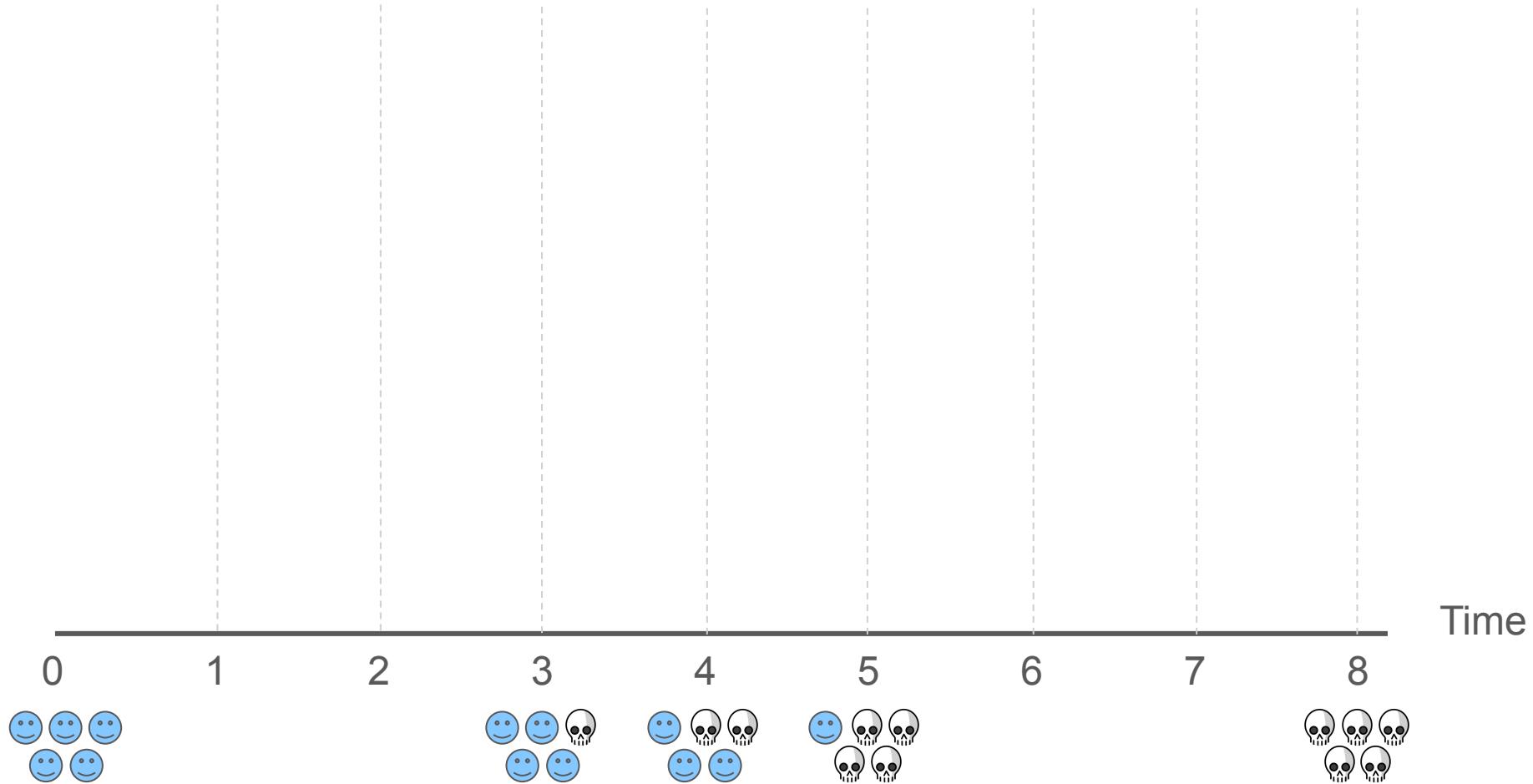
3 years

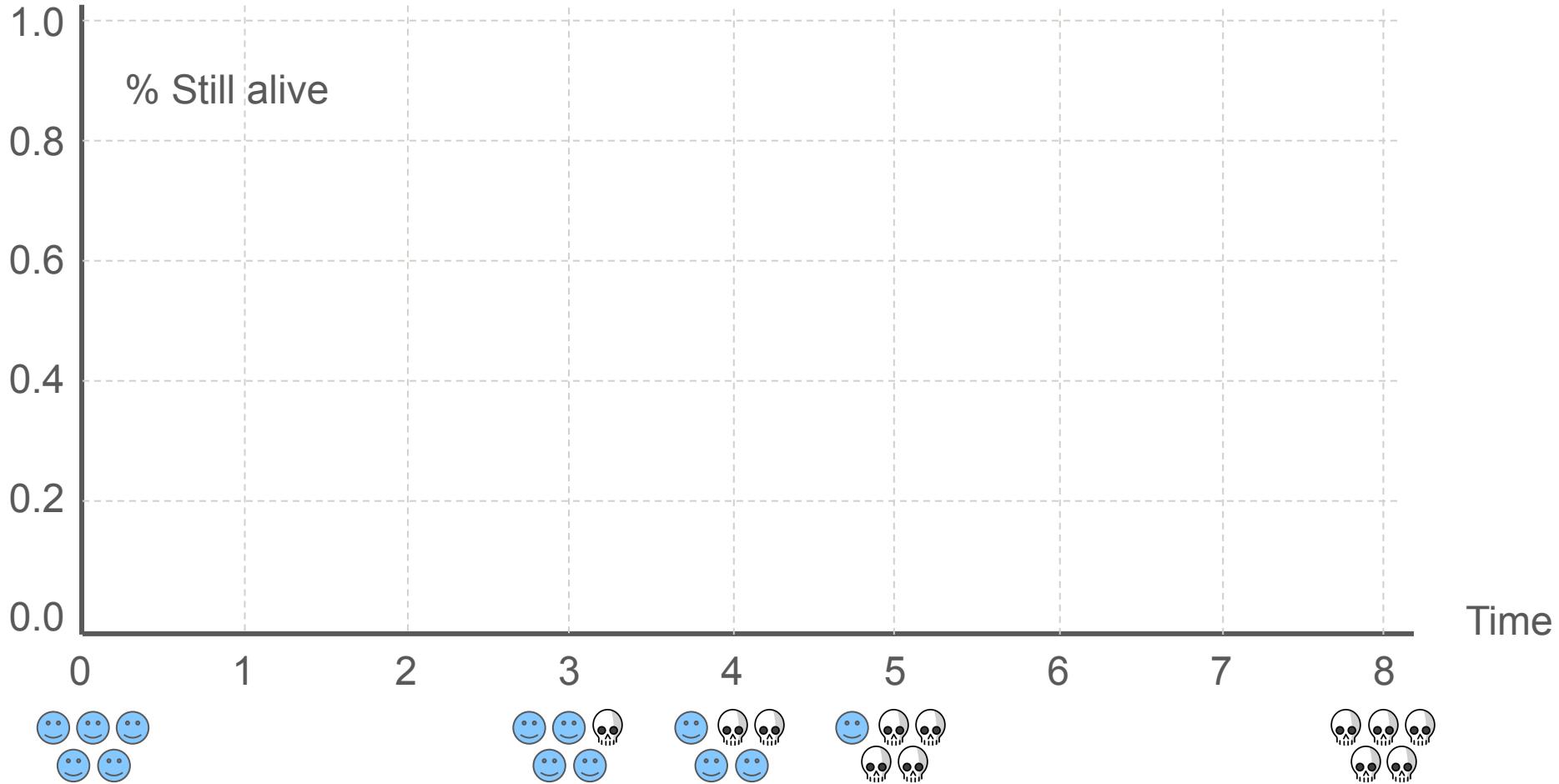


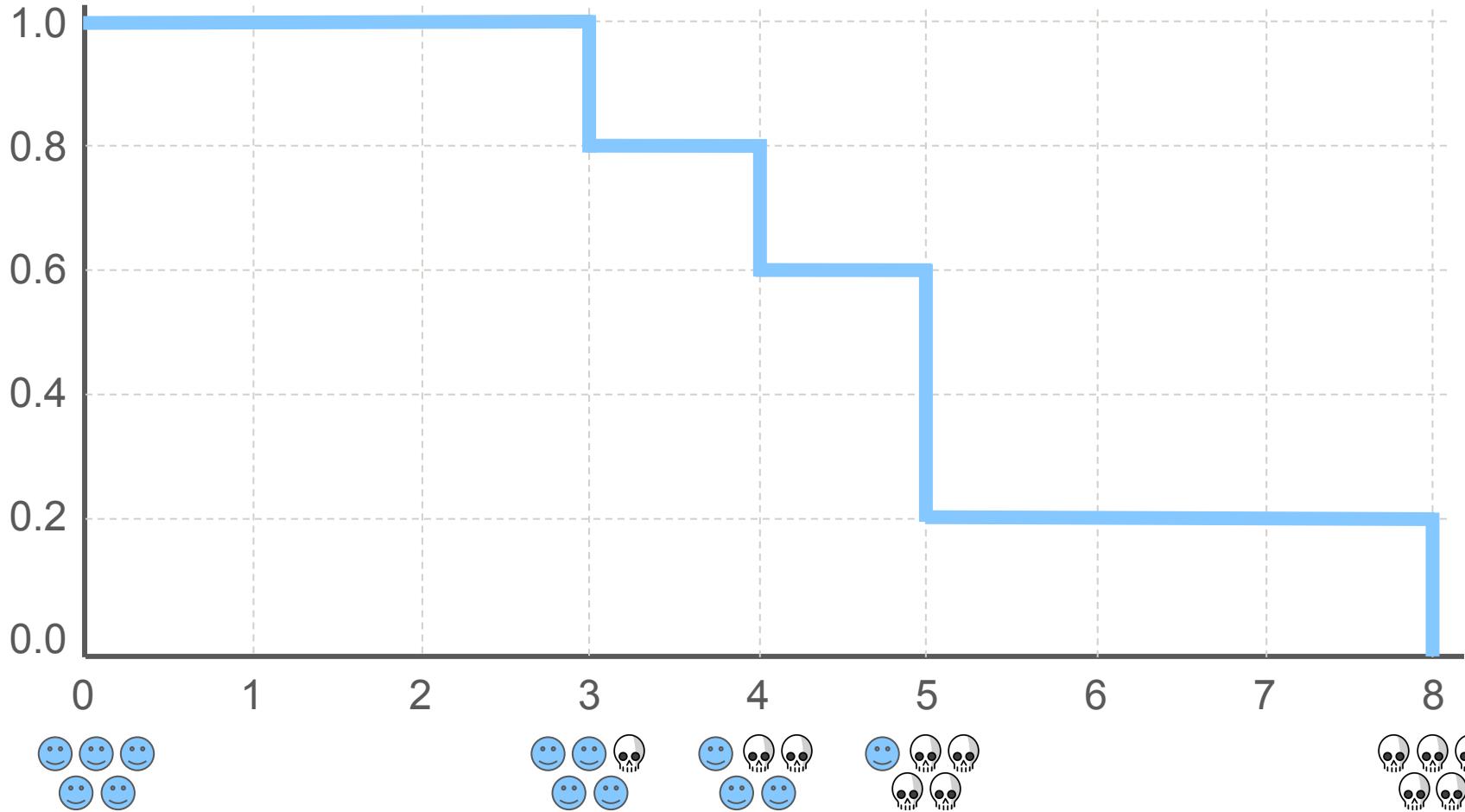
5 years

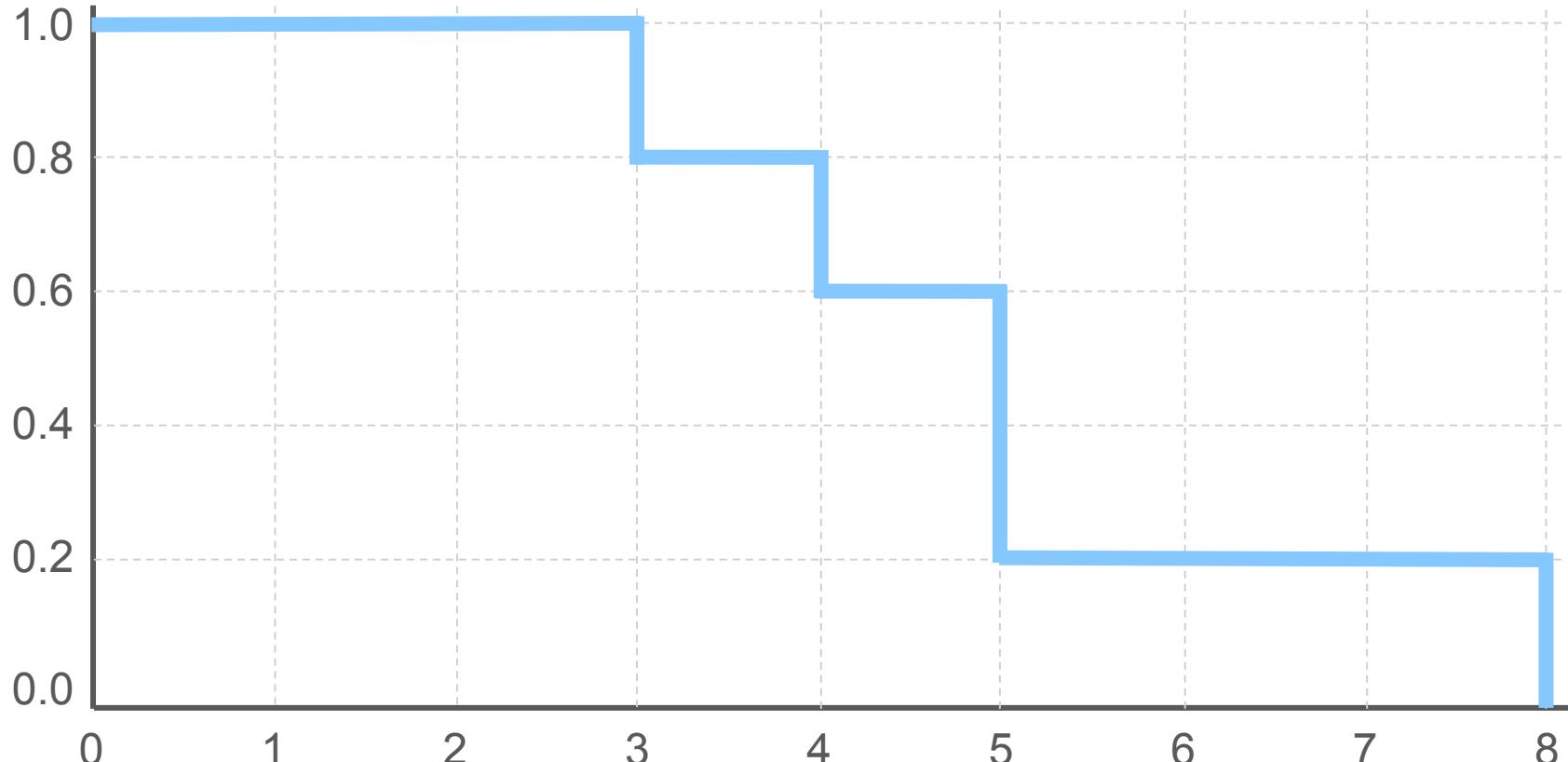
5 years



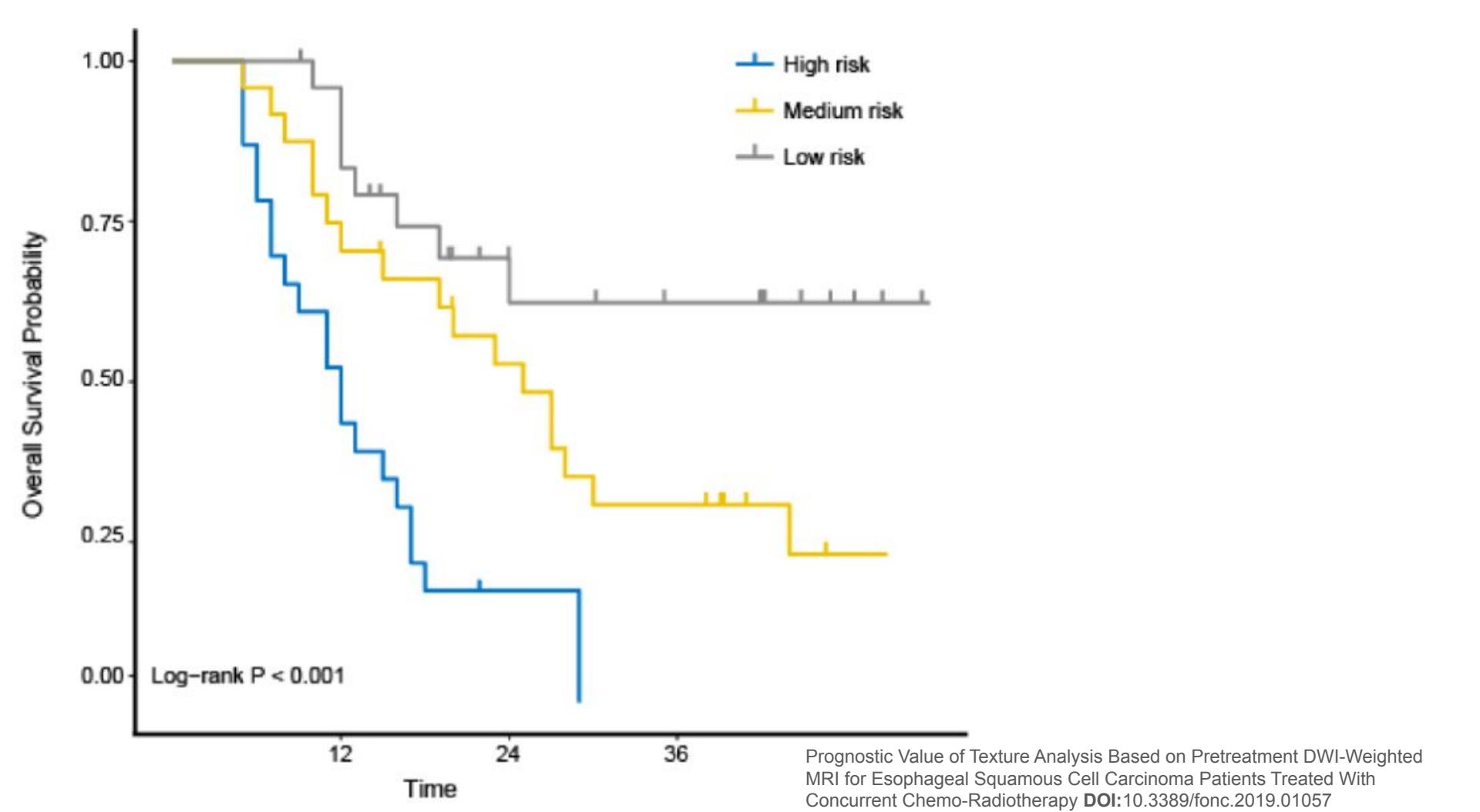


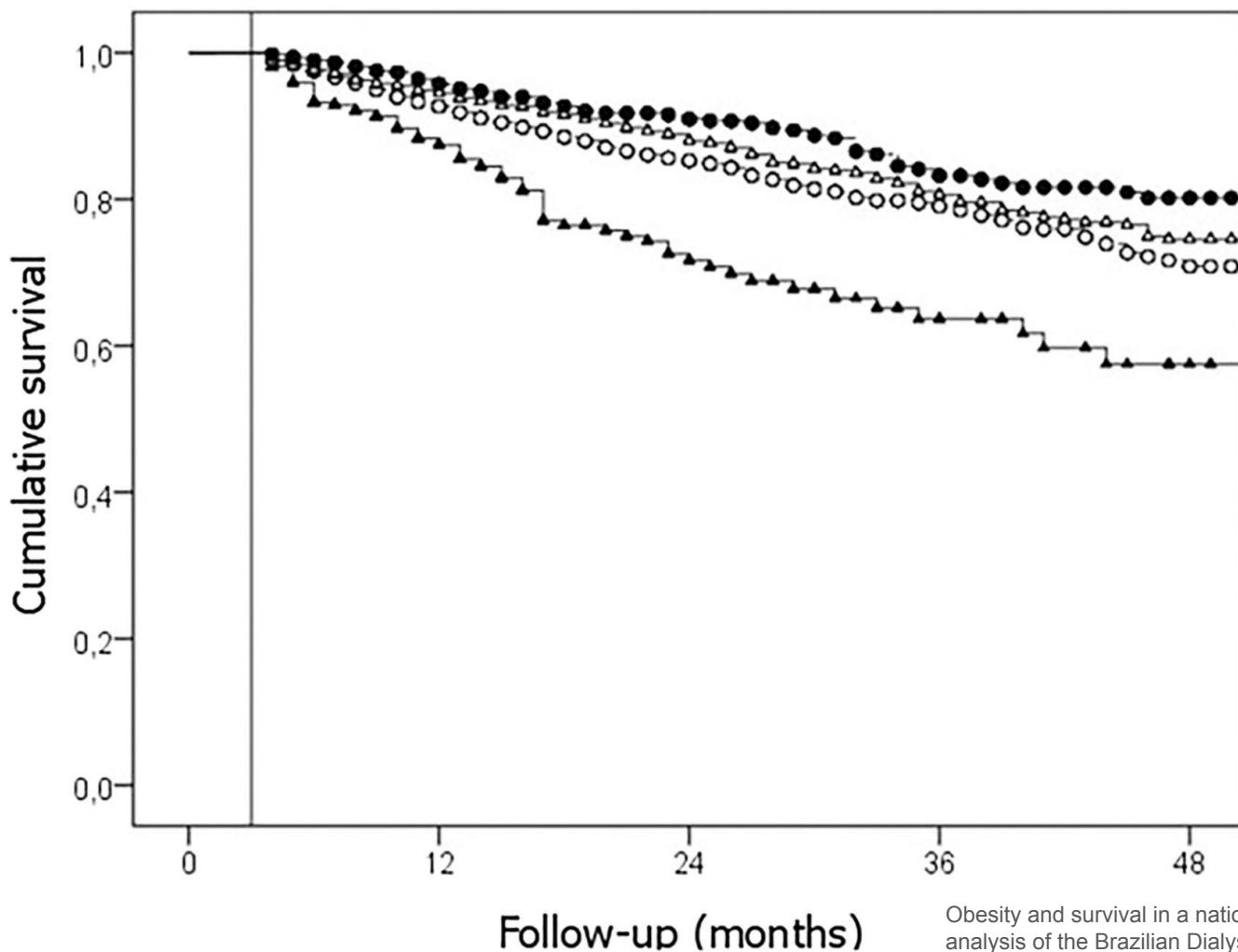




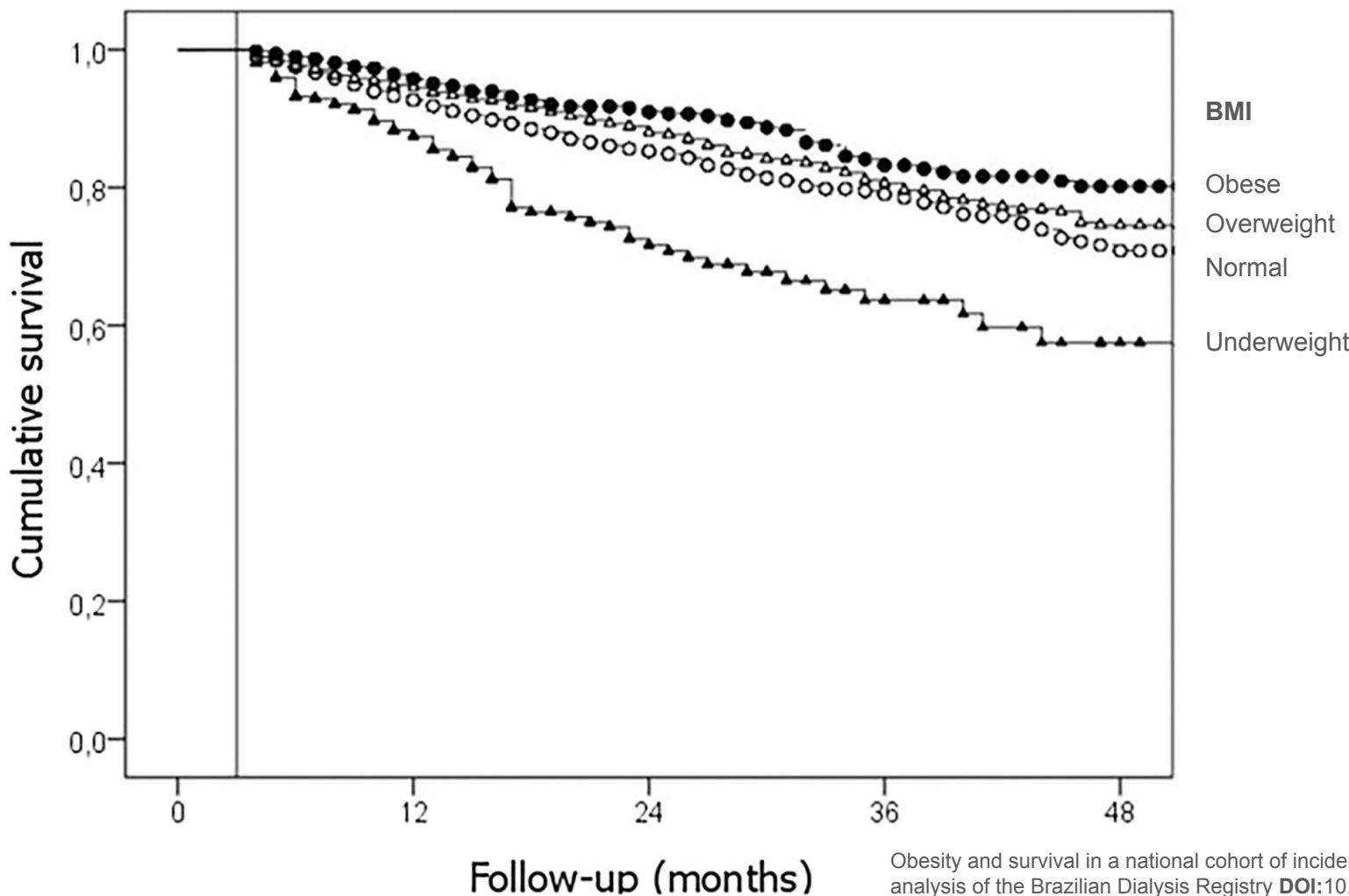


Kaplan-Meier curve





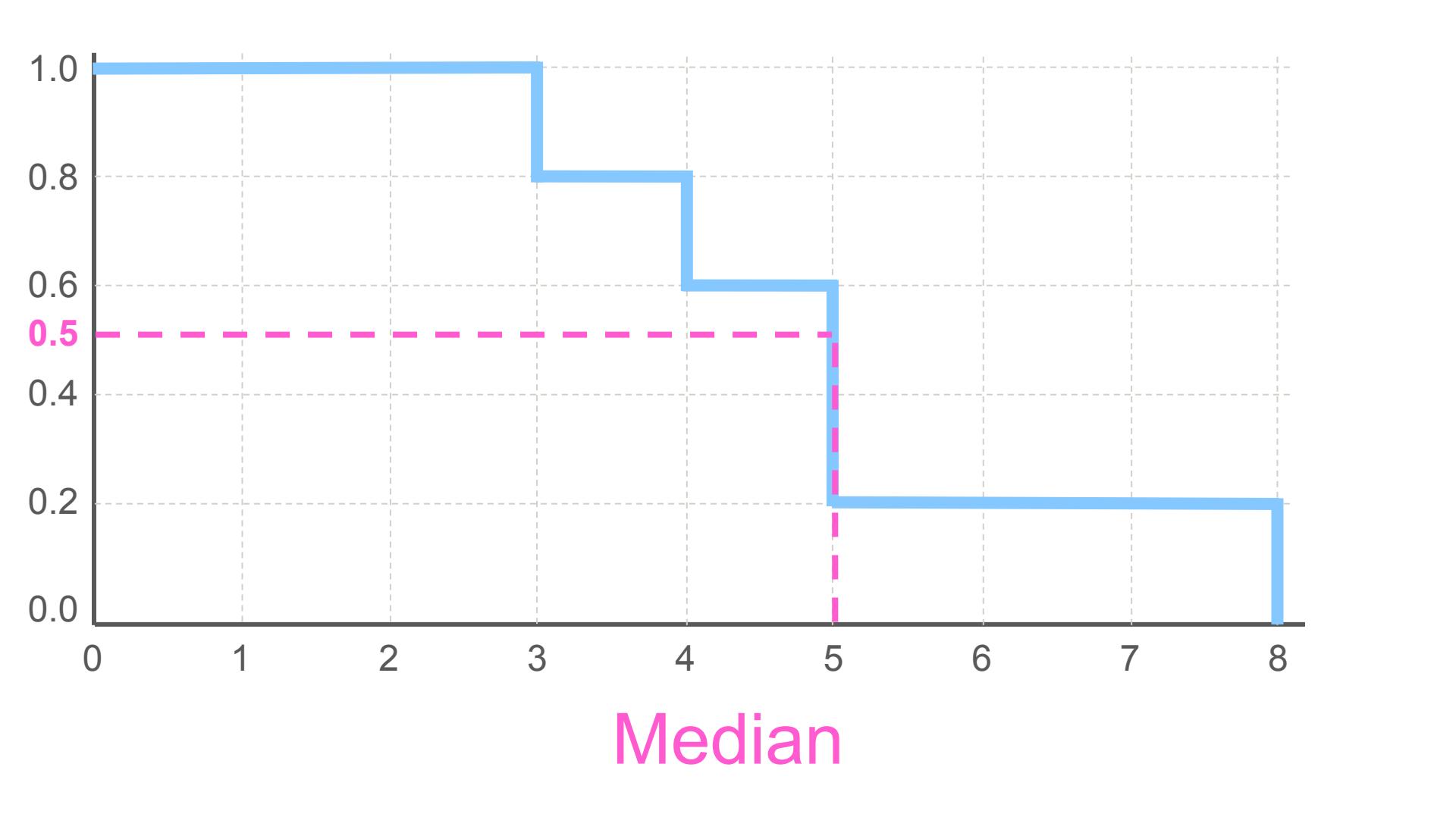
Obesity and survival in a national cohort of incident hemodialysis patients: An analysis of the Brazilian Dialysis Registry DOI:10.1111/hdi.13099

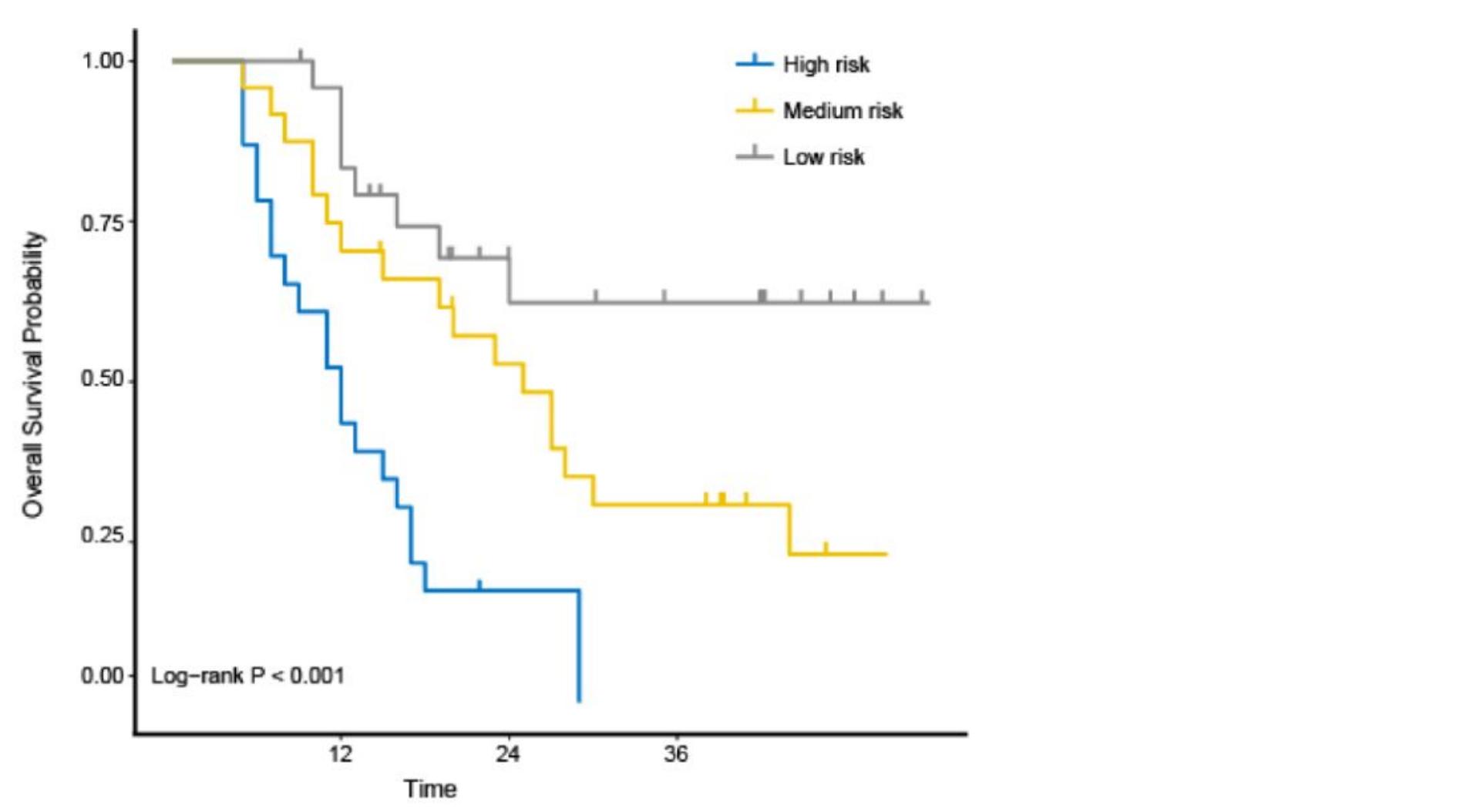


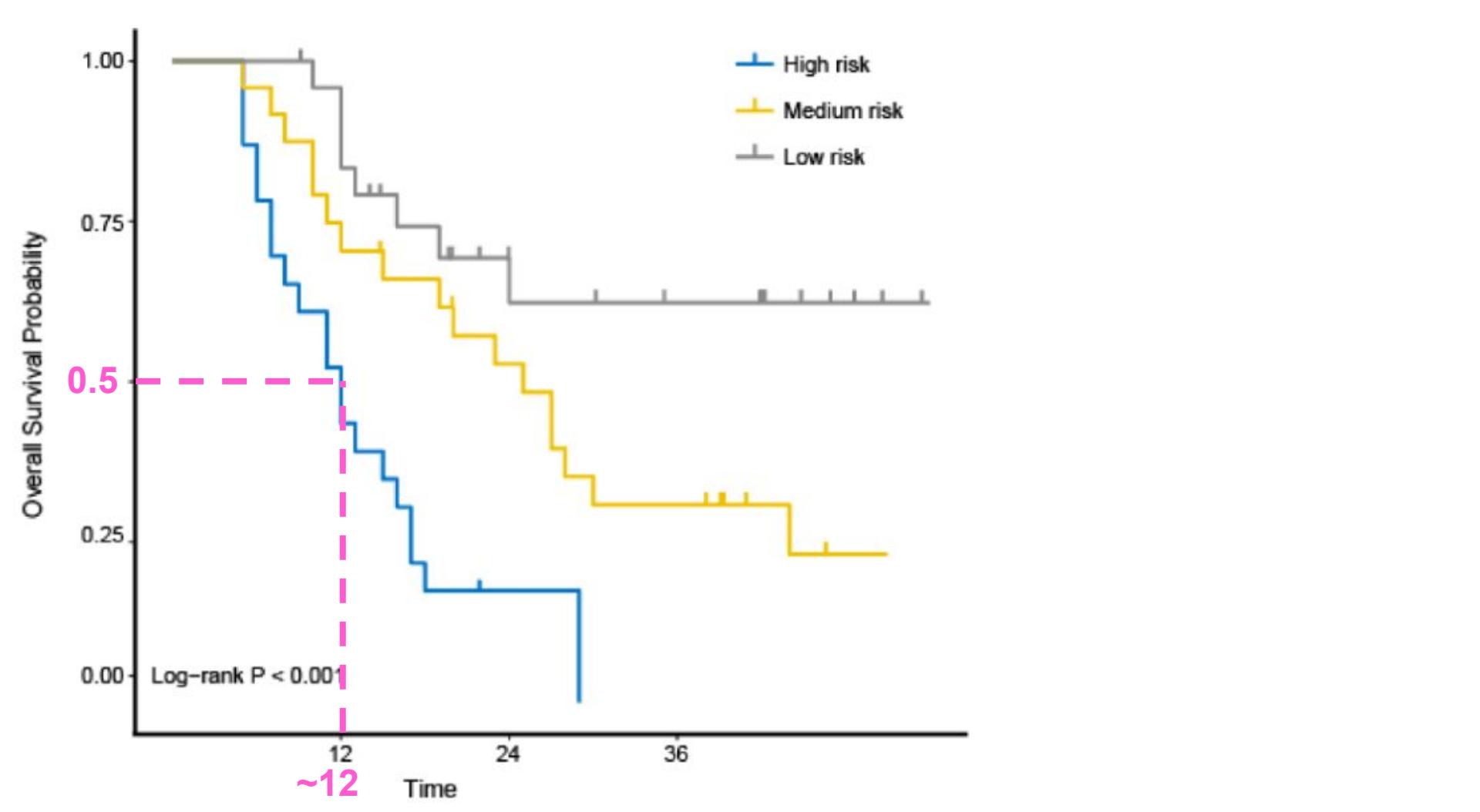
Obesity and survival in a national cohort of incident hemodialysis patients: An analysis of the Brazilian Dialysis Registry DOI:10.1111/hdi.13099

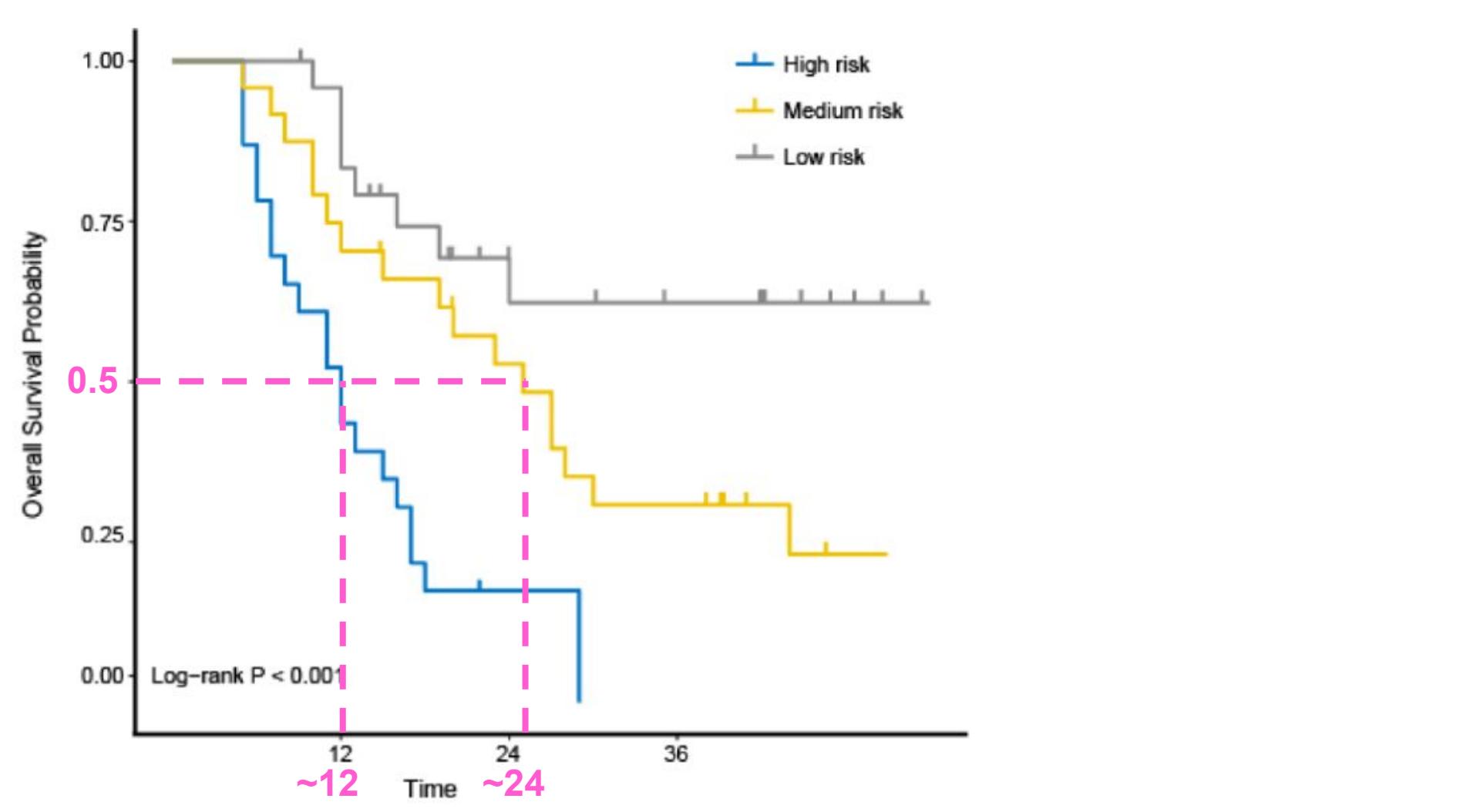
Finding the Mean / Median

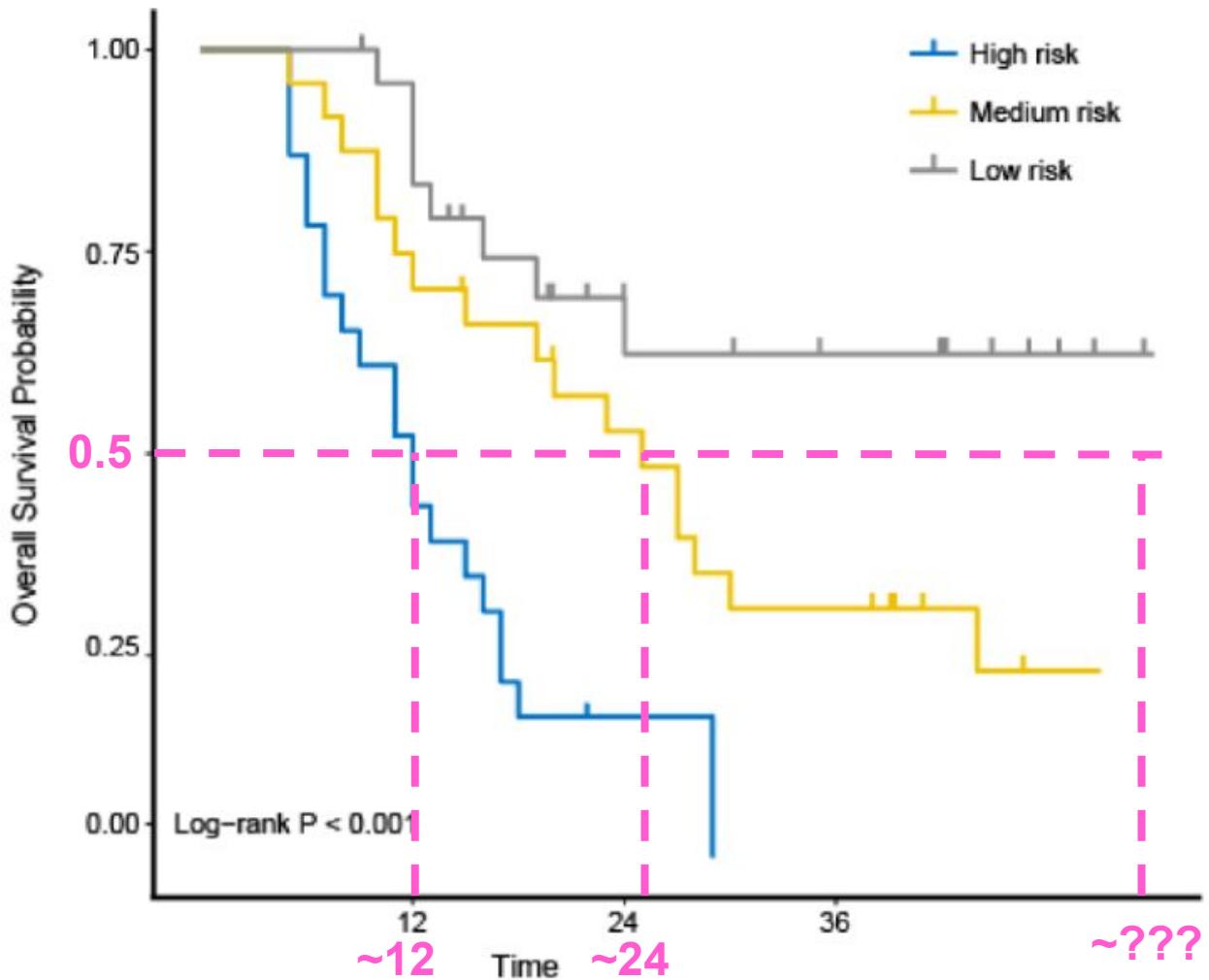
Variance

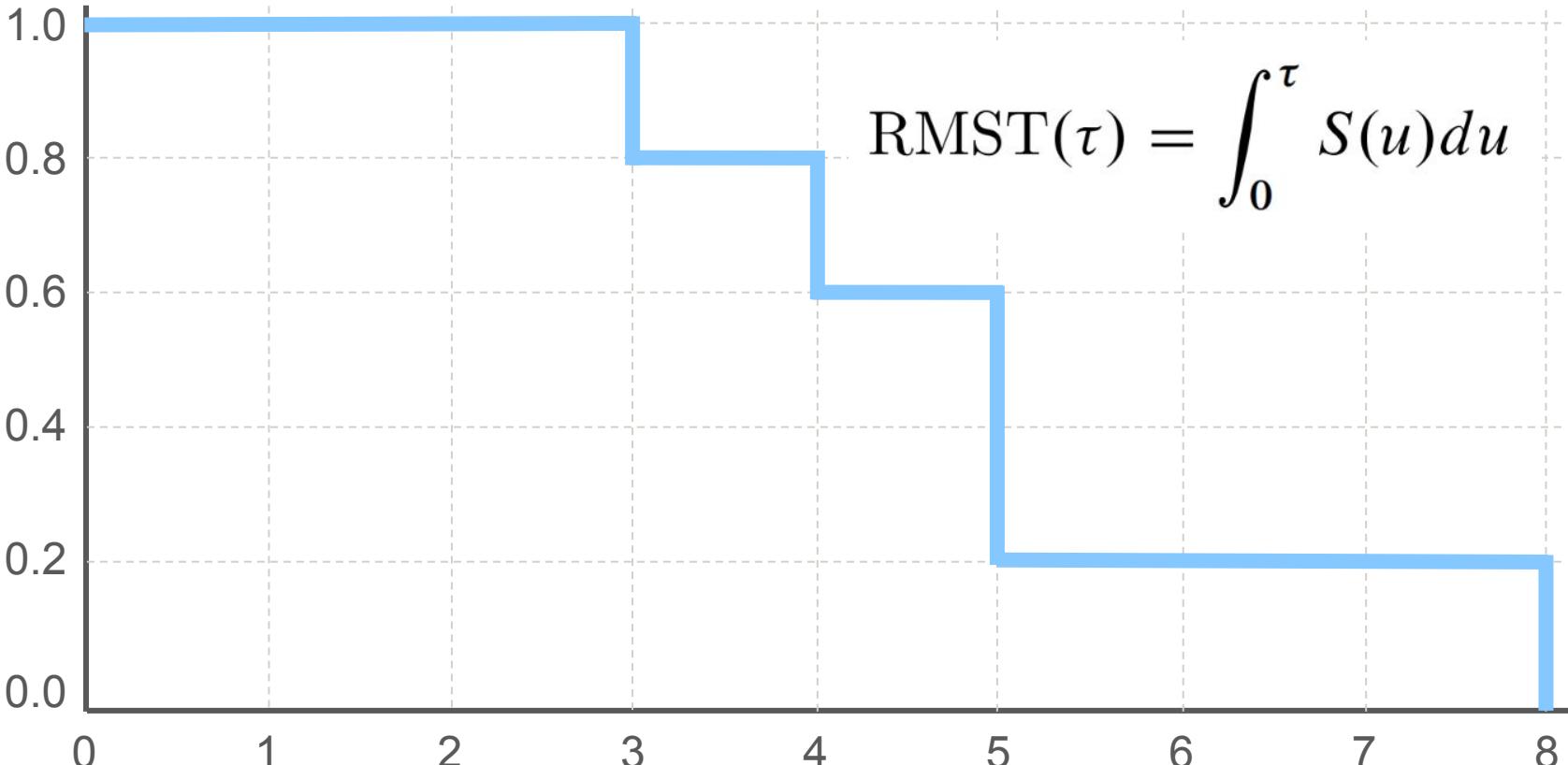




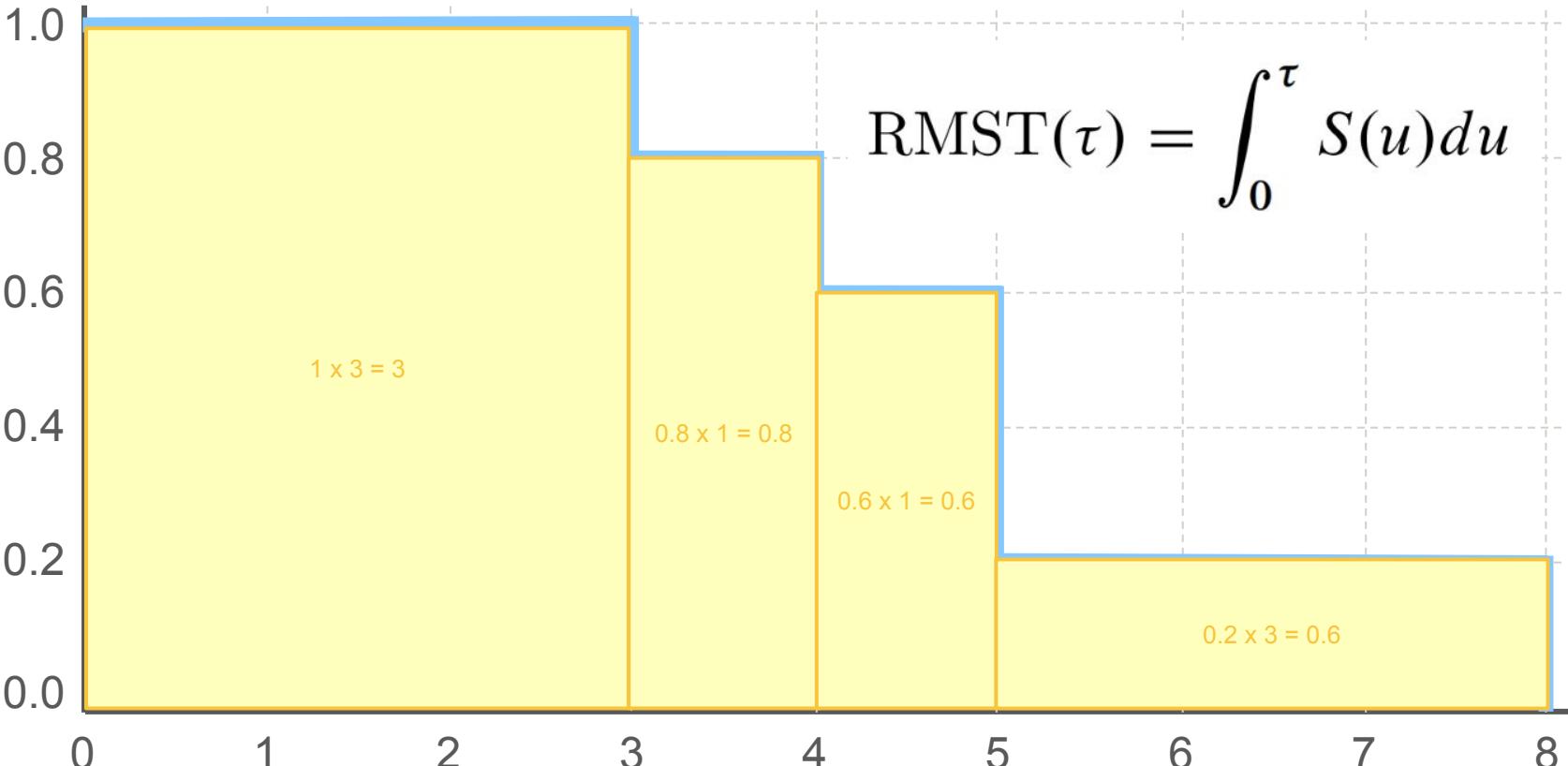




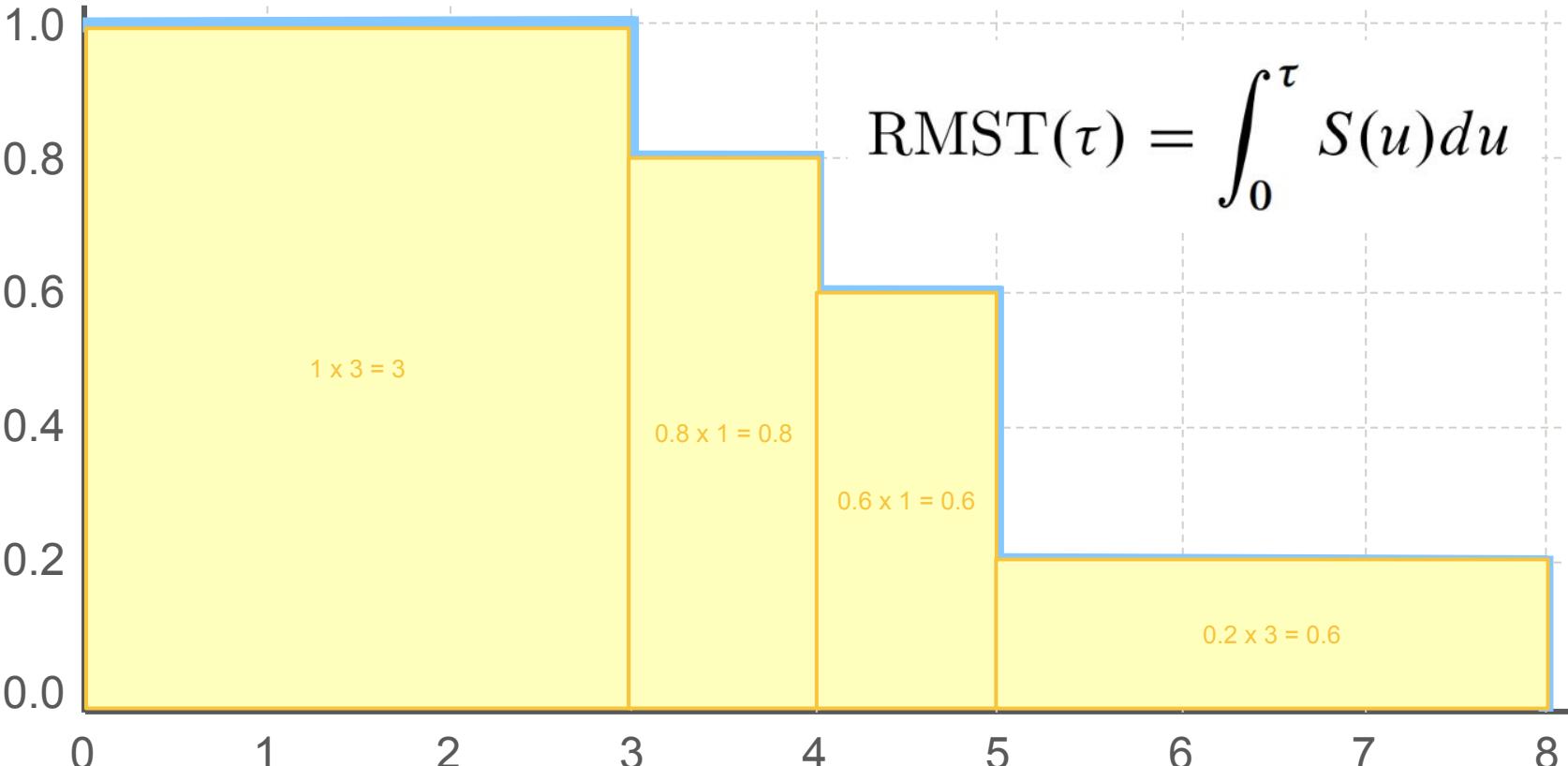




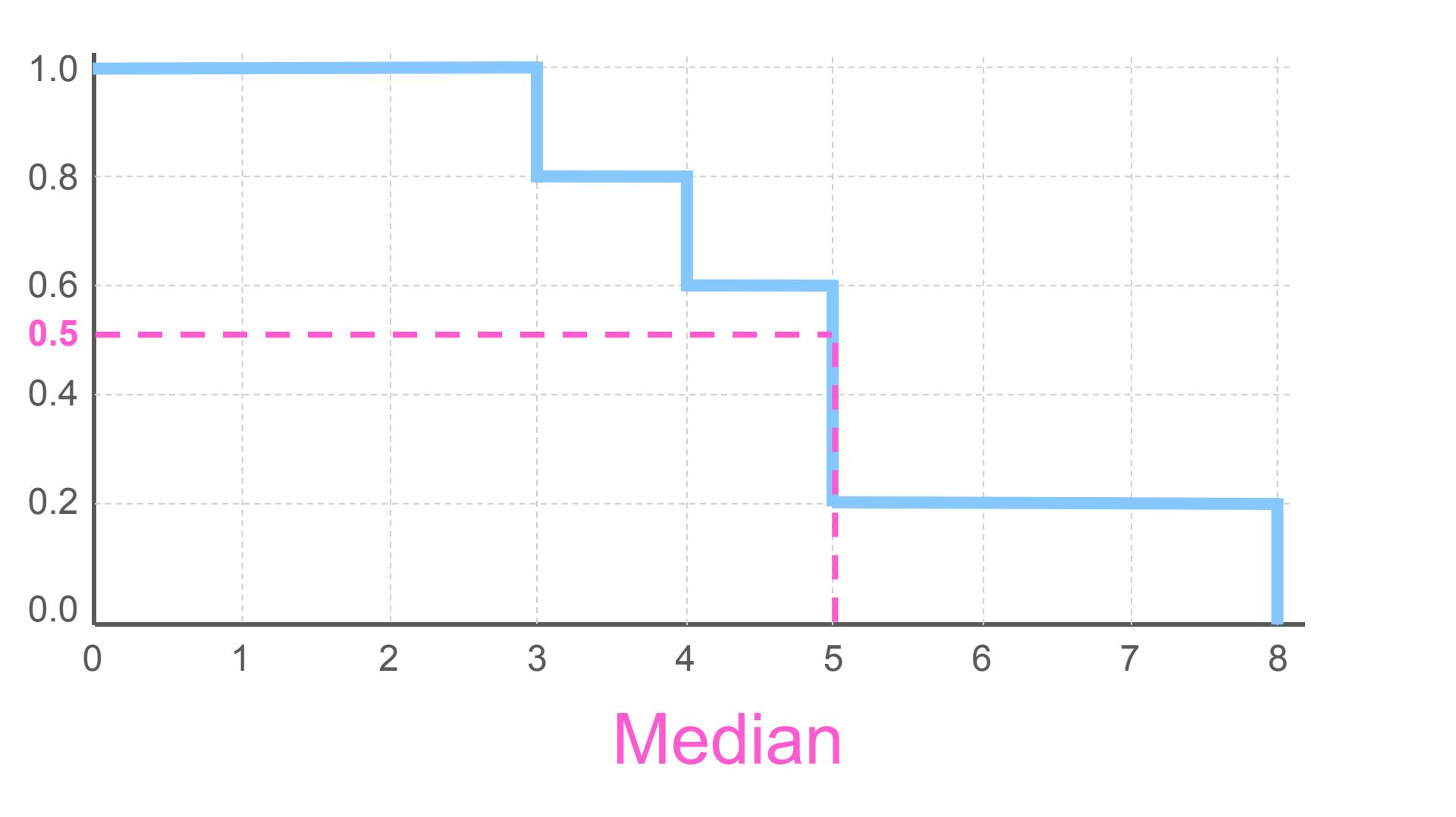
Restricted Mean Survival Time (RMST)

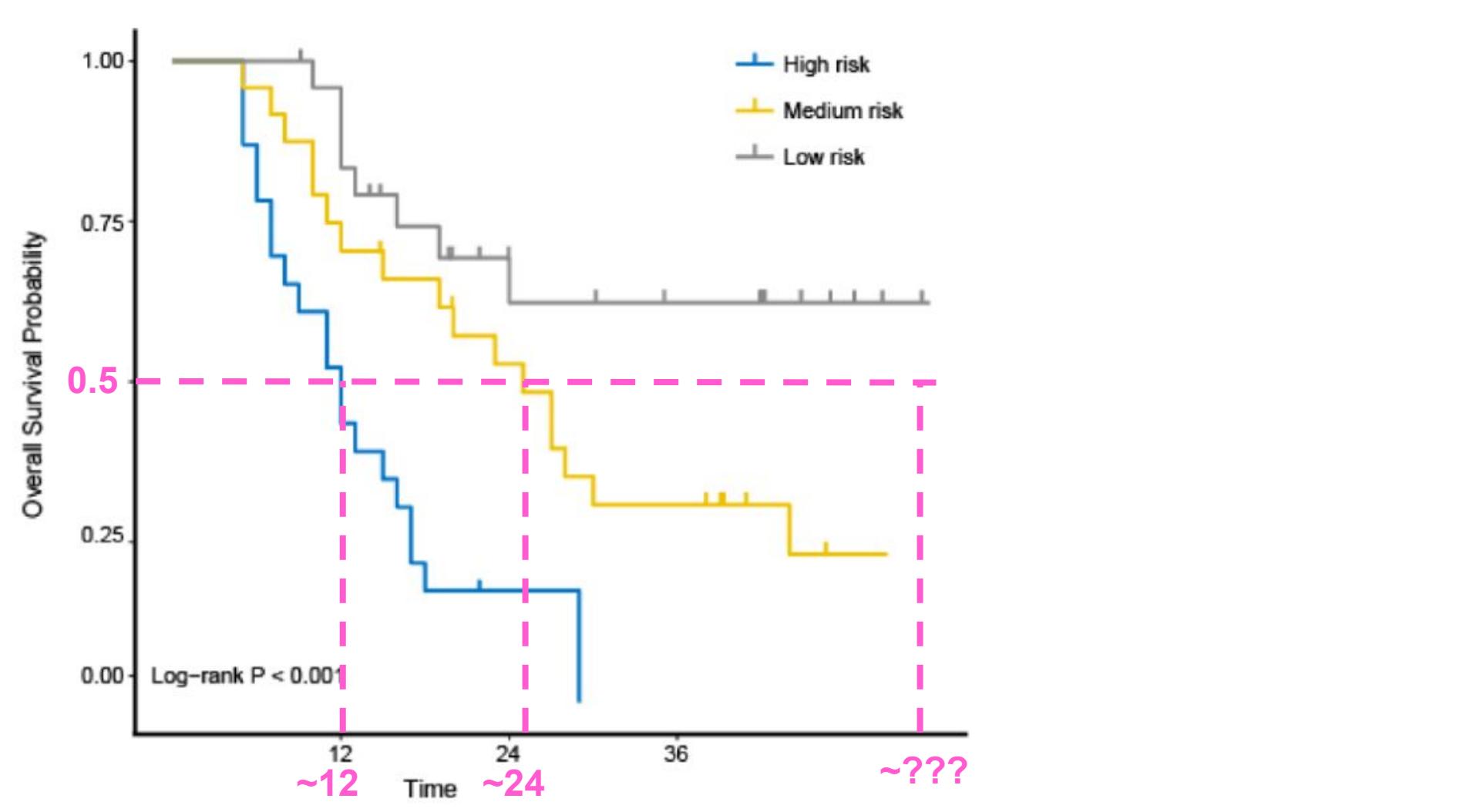


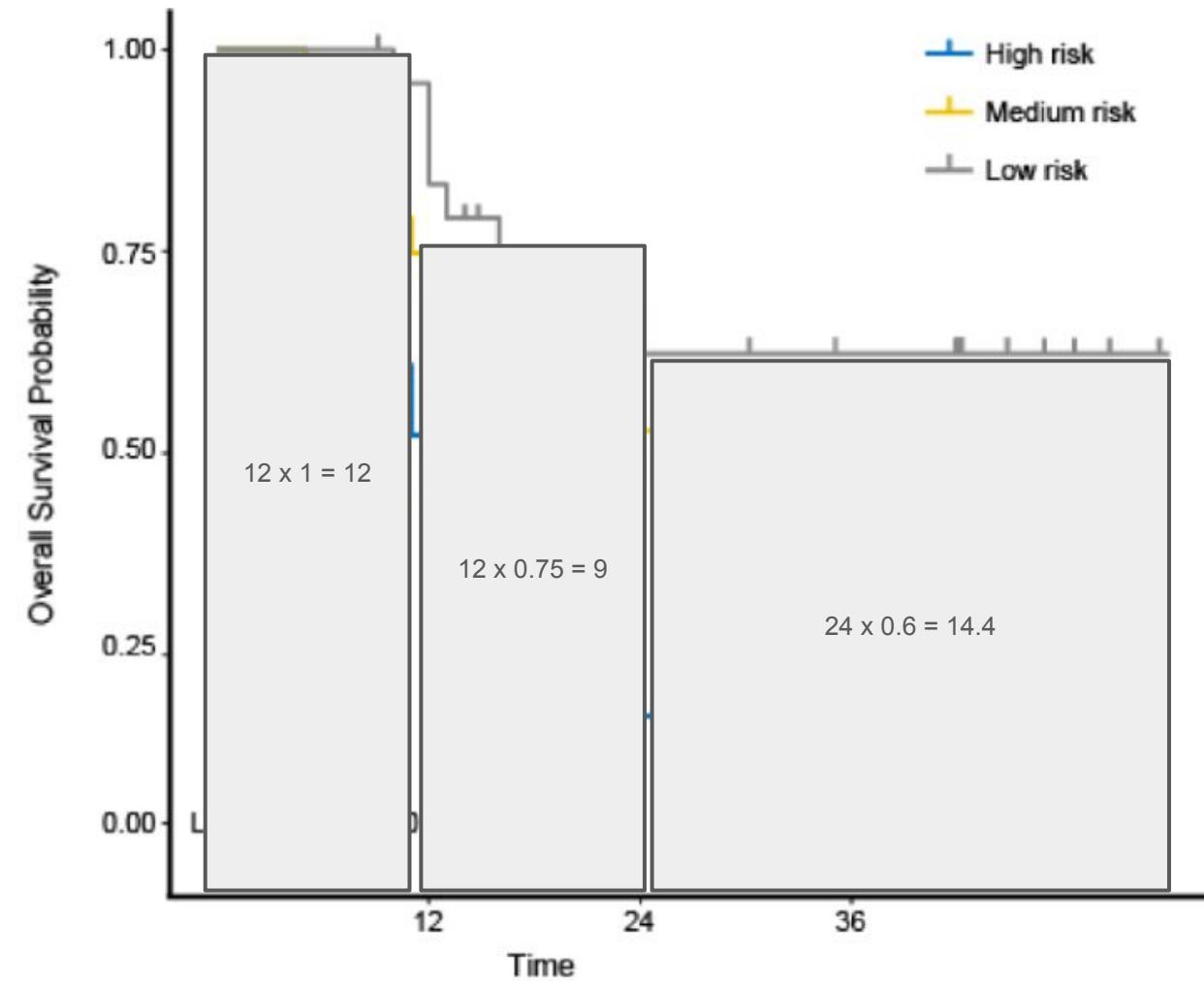
Restricted Mean Survival Time (RMST)

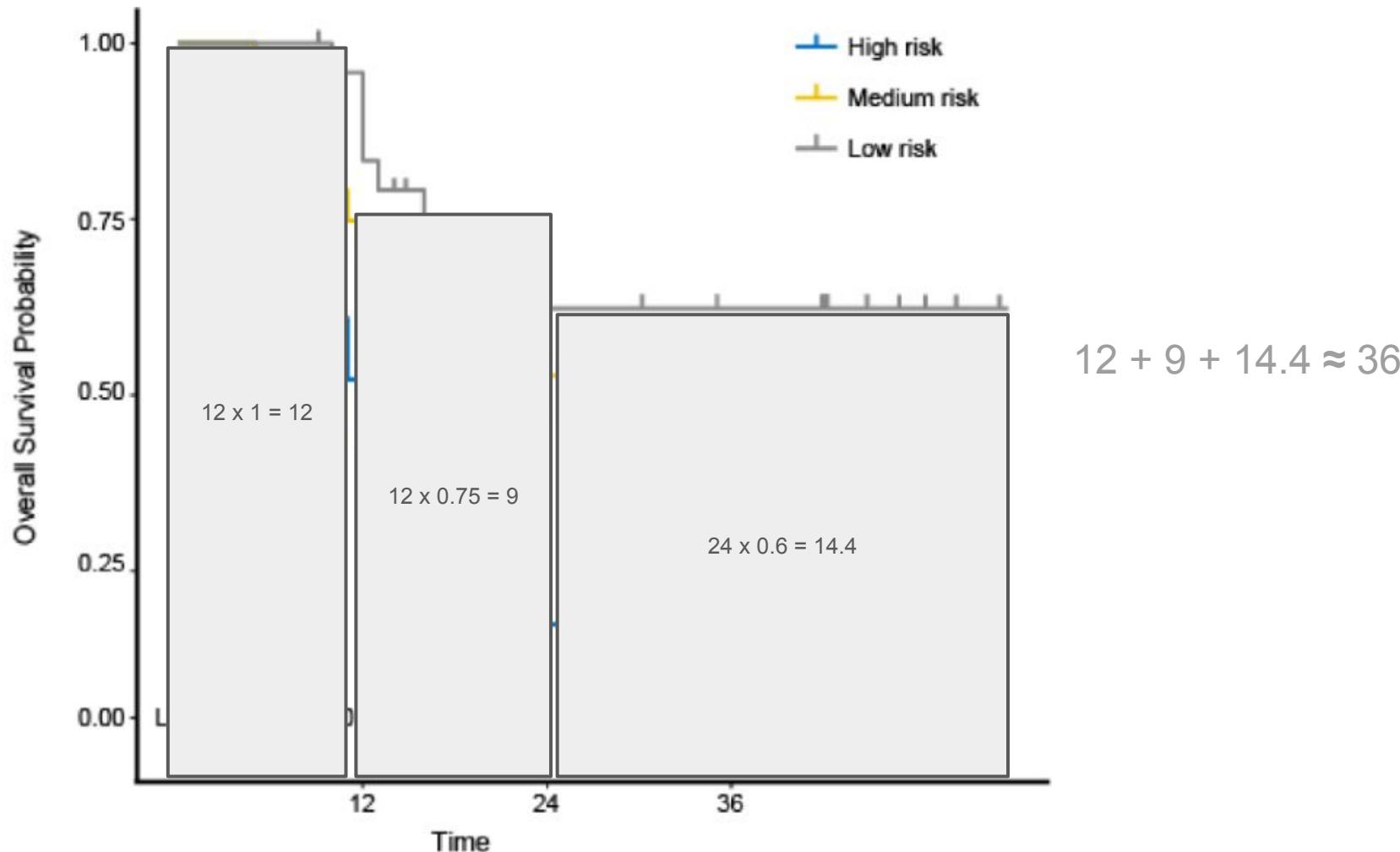


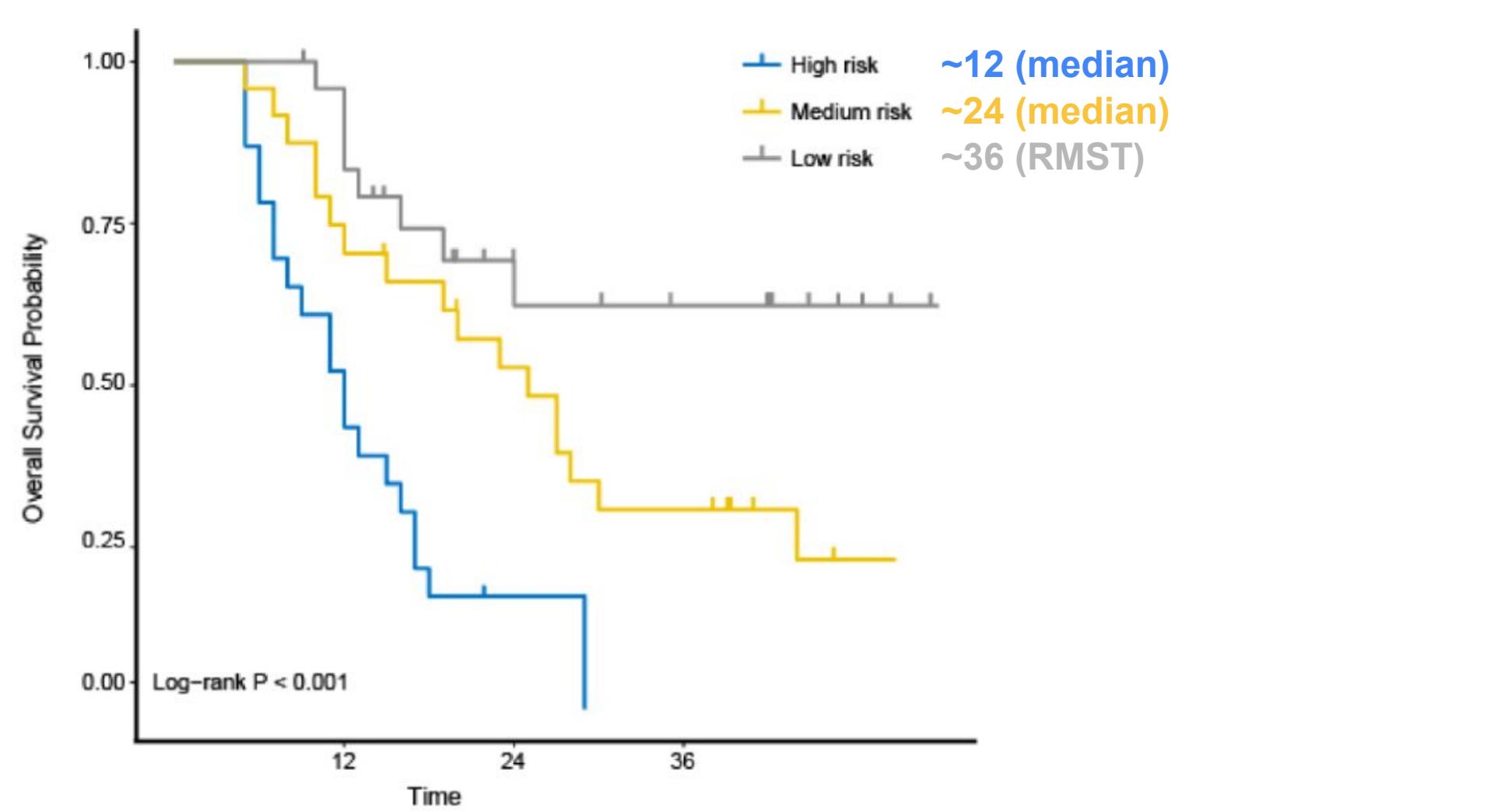
$$3 + 0.8 + 0.6 + 0.6 = 5$$

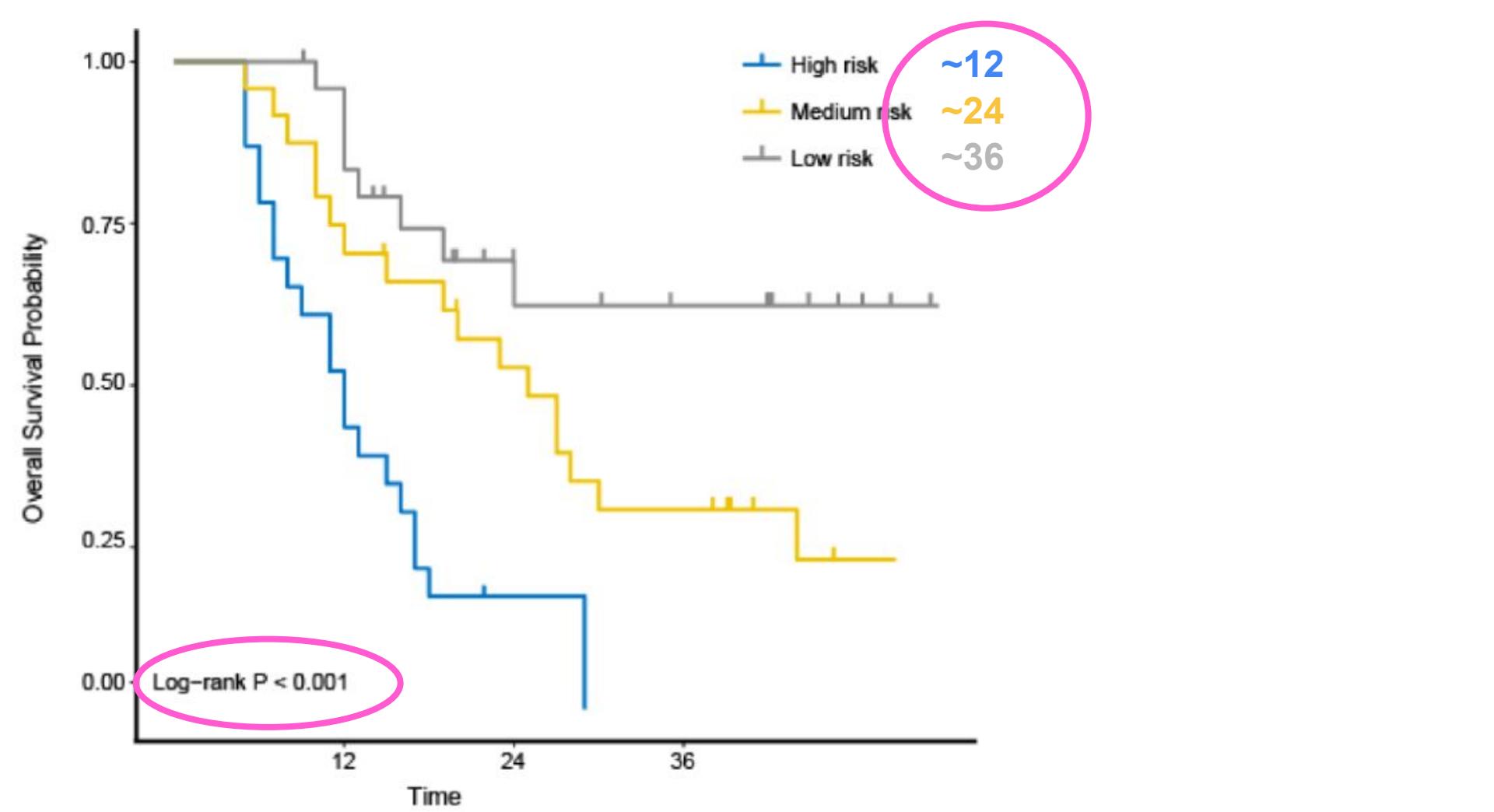


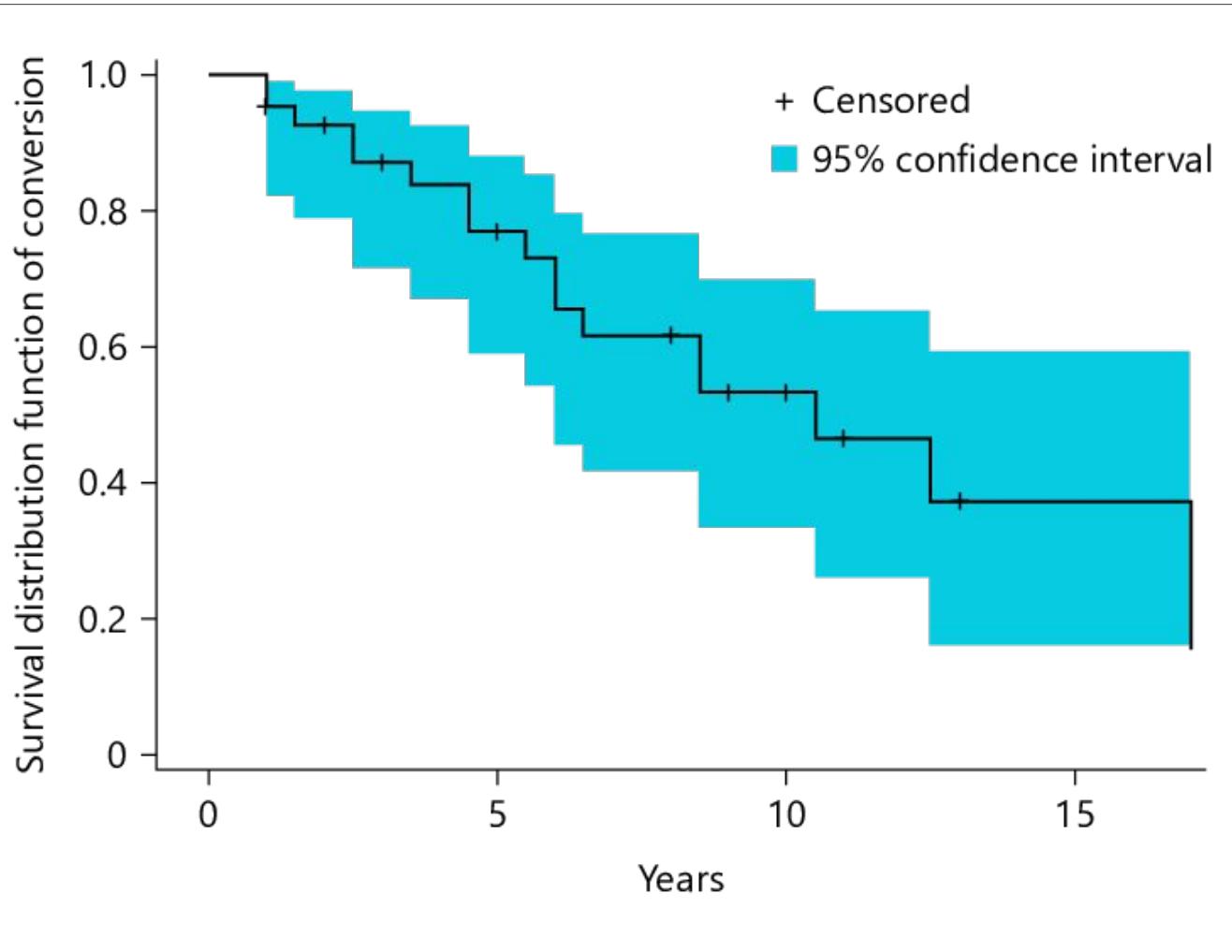












Local Allergic Nasal
Reactions Convert to
Classic Systemic Allergic
Reactions: A Long-Term
Follow-Up

DOI:10.1159/000380852

$$CI = \bar{x} \pm z \frac{s}{\sqrt{n}}$$

CI = confidence interval

\bar{x} = sample mean

z = confidence level value

s = sample standard deviation

n = sample size

5 [3.5 , 6.5]

$$CI = \bar{x} \pm z \frac{s}{\sqrt{n}}$$

CI = confidence interval

s = sample standard deviation

\bar{x} = sample mean

n = sample size

z = confidence level value

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}},$$

$$CI = \bar{x} \pm z \frac{s}{\sqrt{n}}$$

CI = confidence interval

\bar{x} = sample mean

z = confidence level value

s = sample standard deviation

n = sample size

S Approximation

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}},$$

$$CI = \bar{x} \pm z \frac{s}{\sqrt{n}}$$

CI = confidence interval

\bar{x} = sample mean

z = confidence level value

s = sample standard deviation

n = sample size

Greenwood's formula

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}},$$

$$CI = \bar{x} \pm z \frac{s}{\sqrt{n}}$$

CI = confidence interval

s = sample standard deviation

\bar{x} = sample mean

n = sample size

z = confidence level value

Greenwood's formula

$$\widehat{\text{Var}}\left(\widehat{S}(t)\right) = \widehat{S}(t)^2 \sum_{i: t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}$$

$$CI = \bar{x} \pm z \frac{s}{\sqrt{n}}$$

CI = confidence interval

\bar{x} = sample mean

z = confidence level value

s = sample standard deviation

n = sample size

Greenwood's formula

$$\widehat{\text{Var}}\left(\widehat{S}(t)\right) = \widehat{S}(t)^2 \sum_{i: t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}$$

$$CI = \widehat{S}(t) \pm Z_{1-\alpha/2} \sqrt{\widehat{V}[\widehat{S}(t)]}$$

Greenwood's formula

$$\widehat{\text{Var}}\left(\widehat{S}(t)\right) = \widehat{S}(t)^2 \sum_{i: t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}$$

$$CI = \widehat{S}(t) \pm Z_{1-\alpha/2} \sqrt{\widehat{V}[\widehat{S}(t)]}$$

Survival Time (t)	Risk Before (n)	Events at t (d)	Survival probability S(t)	Lower 95%	Upper 95%
3	5	1	0.8	0.45	1.15
4	4	1	0.6	0.17	1.02
5	3	2	0.2	-0.15	0.55
5	3	2	0.2		

1.0

0.8

0.6

0.4

0.2

0.0

0

1

2

3

4

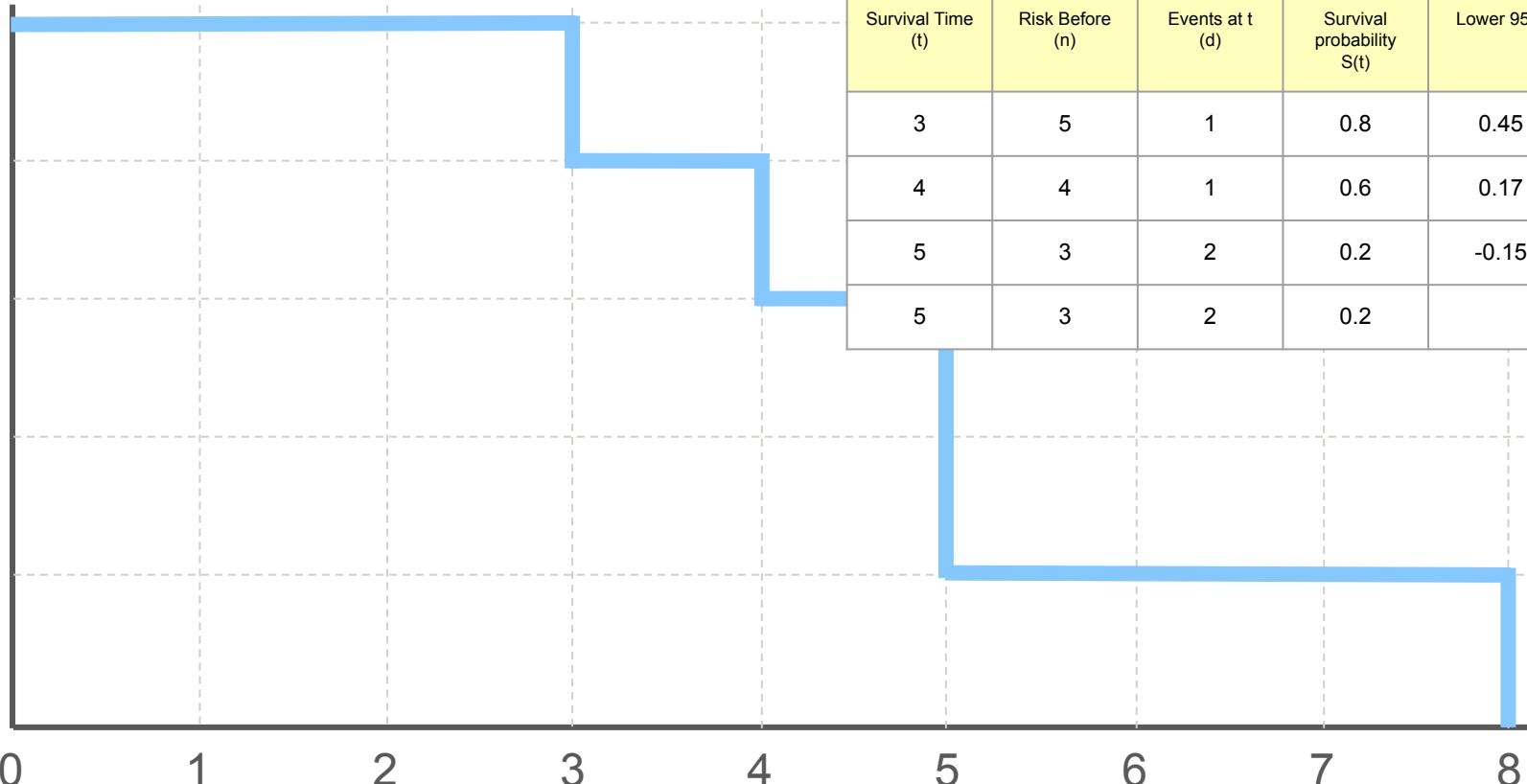
5

6

7

8

Survival Time (t)	Risk Before (n)	Events at t (d)	Survival probability $S(t)$	Lower 95%	Upper 95%
3	5	1	0.8	0.45	1.15
4	4	1	0.6	0.17	1.02
5	3	2	0.2	-0.15	0.55
5	3	2	0.2		



1.0

0.8

0.6

0.4

0.2

0.0

0

1

2

3

4

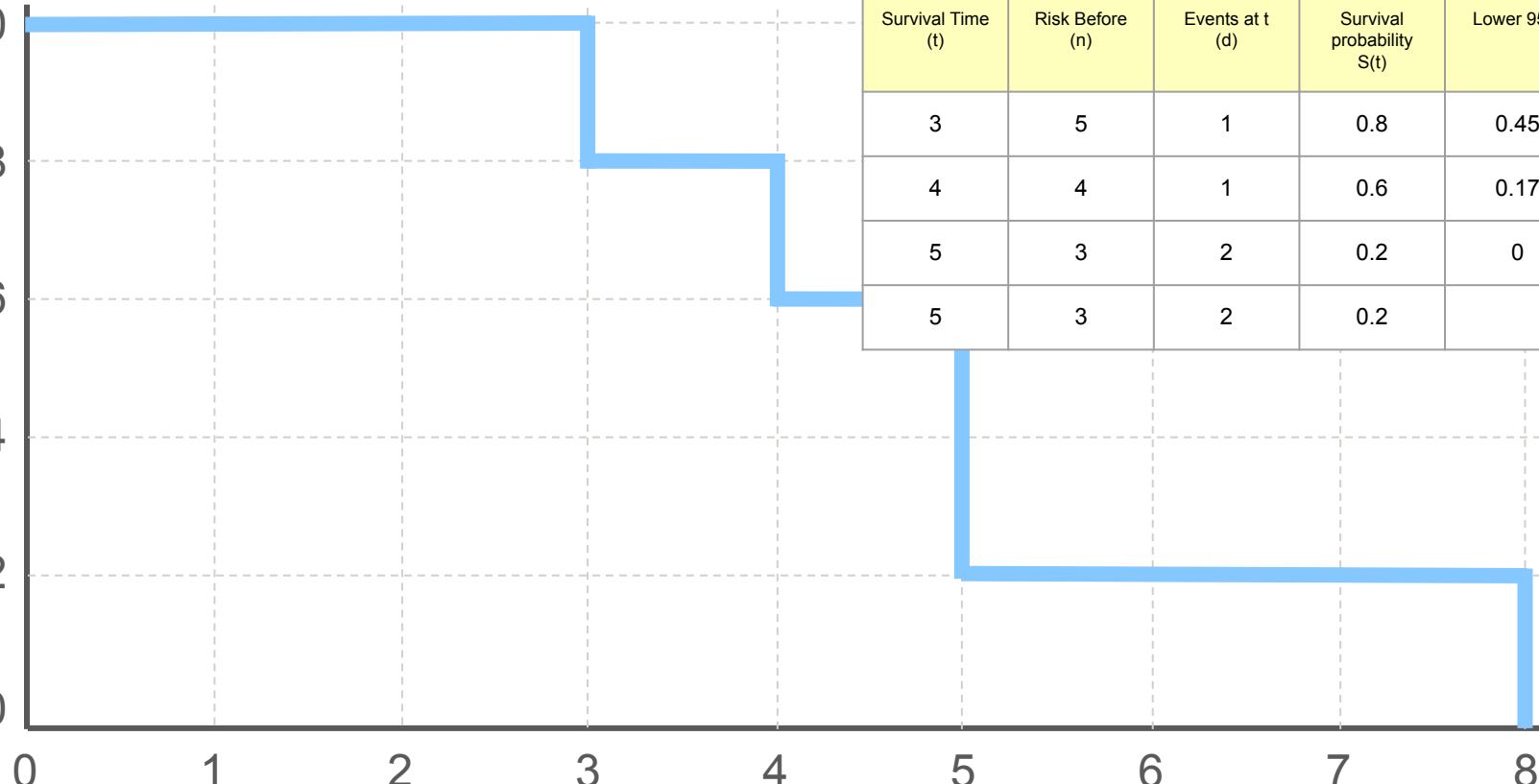
5

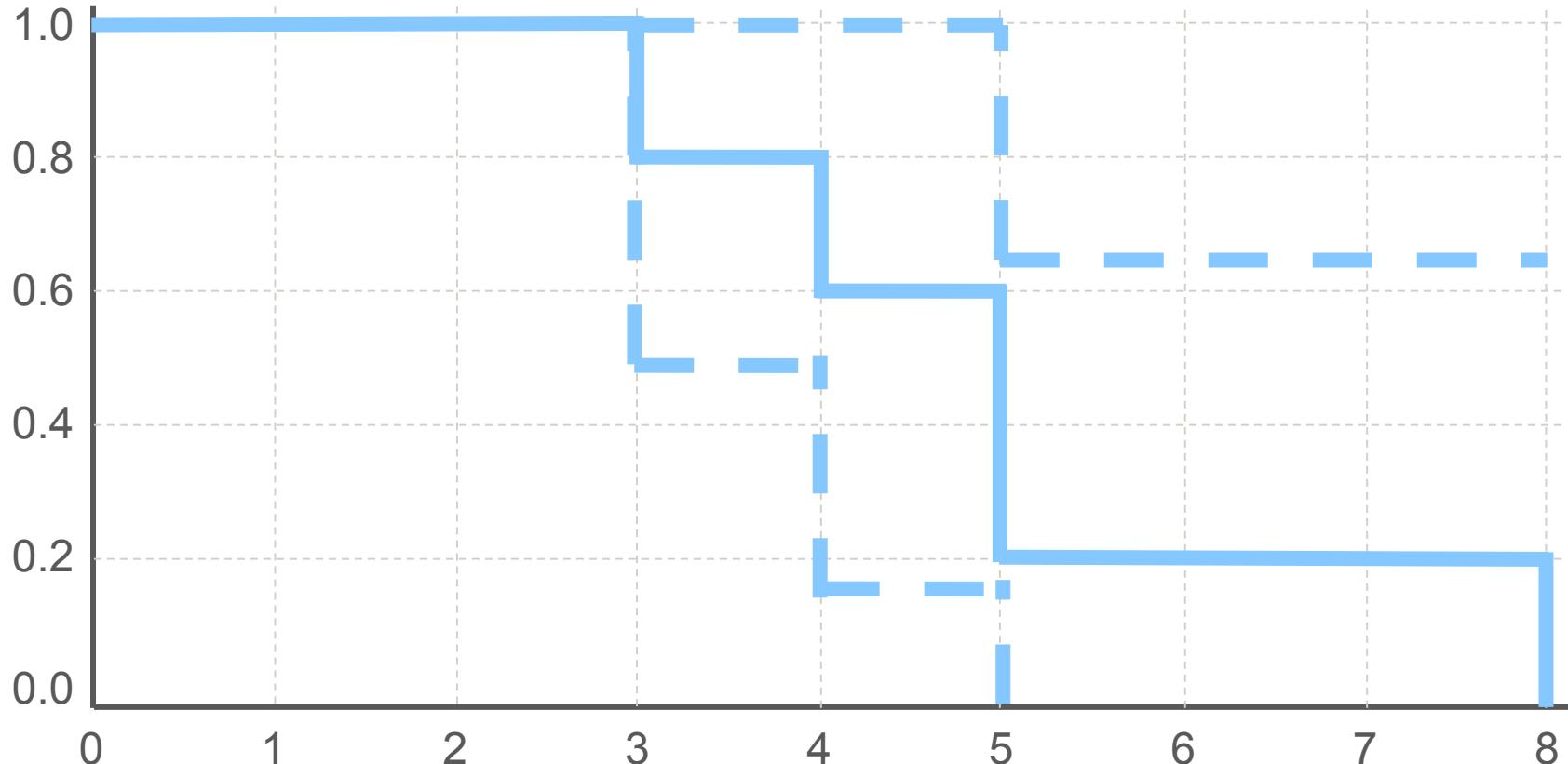
6

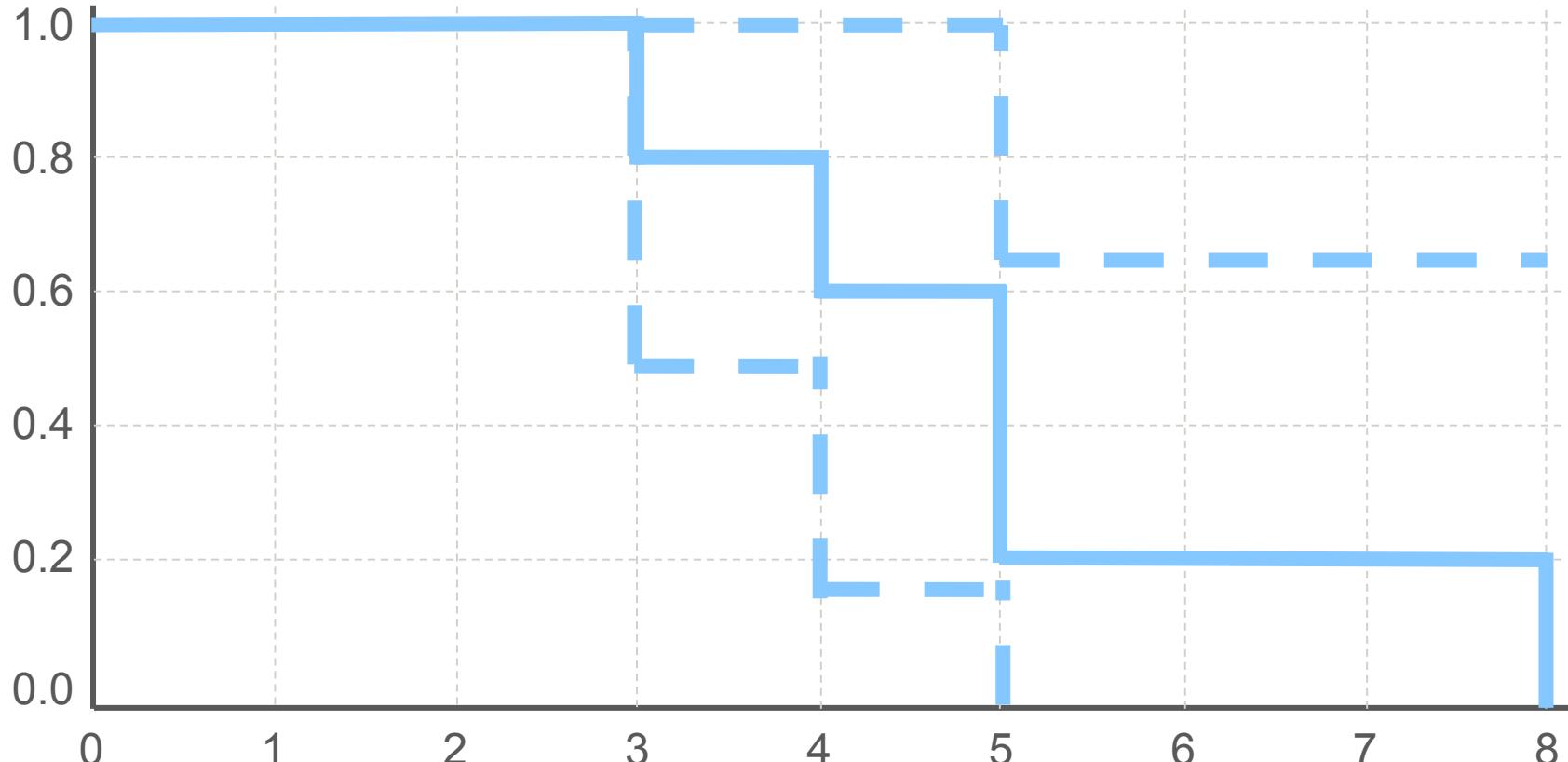
7

8

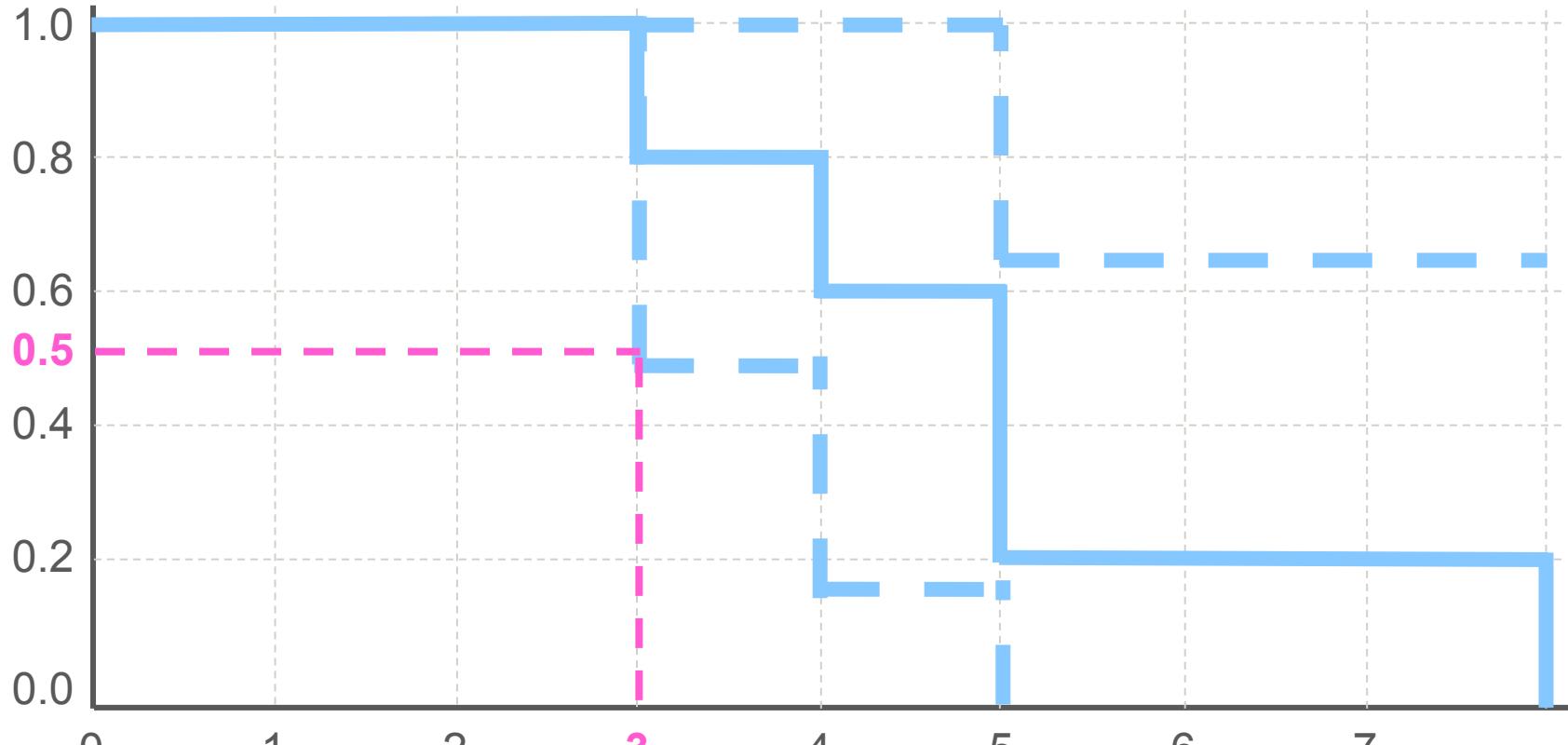
Survival Time (t)	Risk Before (n)	Events at t (d)	Survival probability $S(t)$	Lower 95%	Upper 95%
3	5	1	0.8	0.45	1
4	4	1	0.6	0.17	1
5	3	2	0.2	0	0.55
5	3	2	0.2		



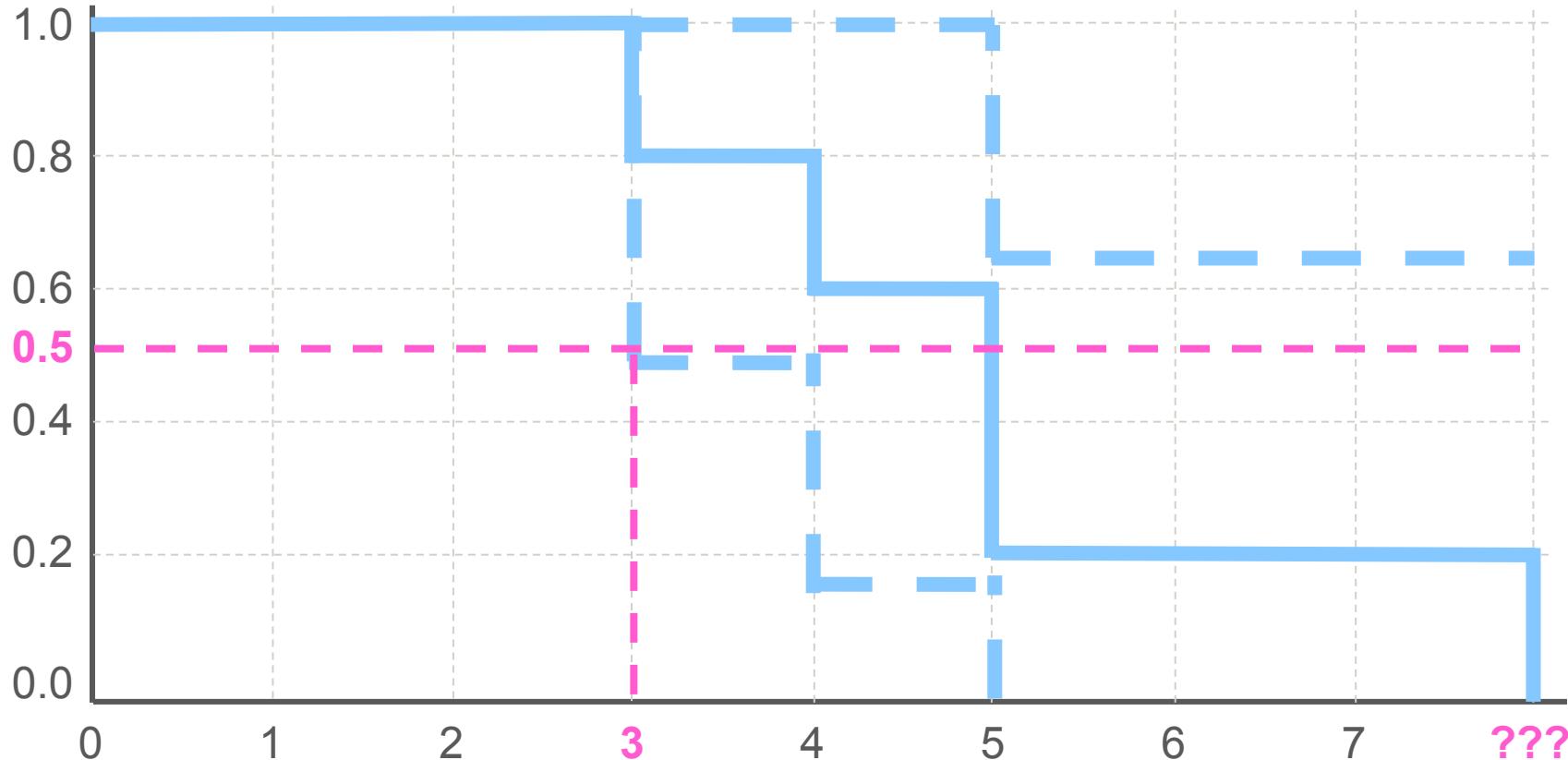




CI 95%



CI 95% Median



CI 95% Median

Greenwood's formula

$$\widehat{\text{Var}}\left(\widehat{S}(t)\right) = \widehat{S}(t)^2 \sum_{i: t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}$$

$$CI = \widehat{S}(t) \pm Z_{1-\alpha/2} \sqrt{\widehat{V}[\widehat{S}(t)]}$$

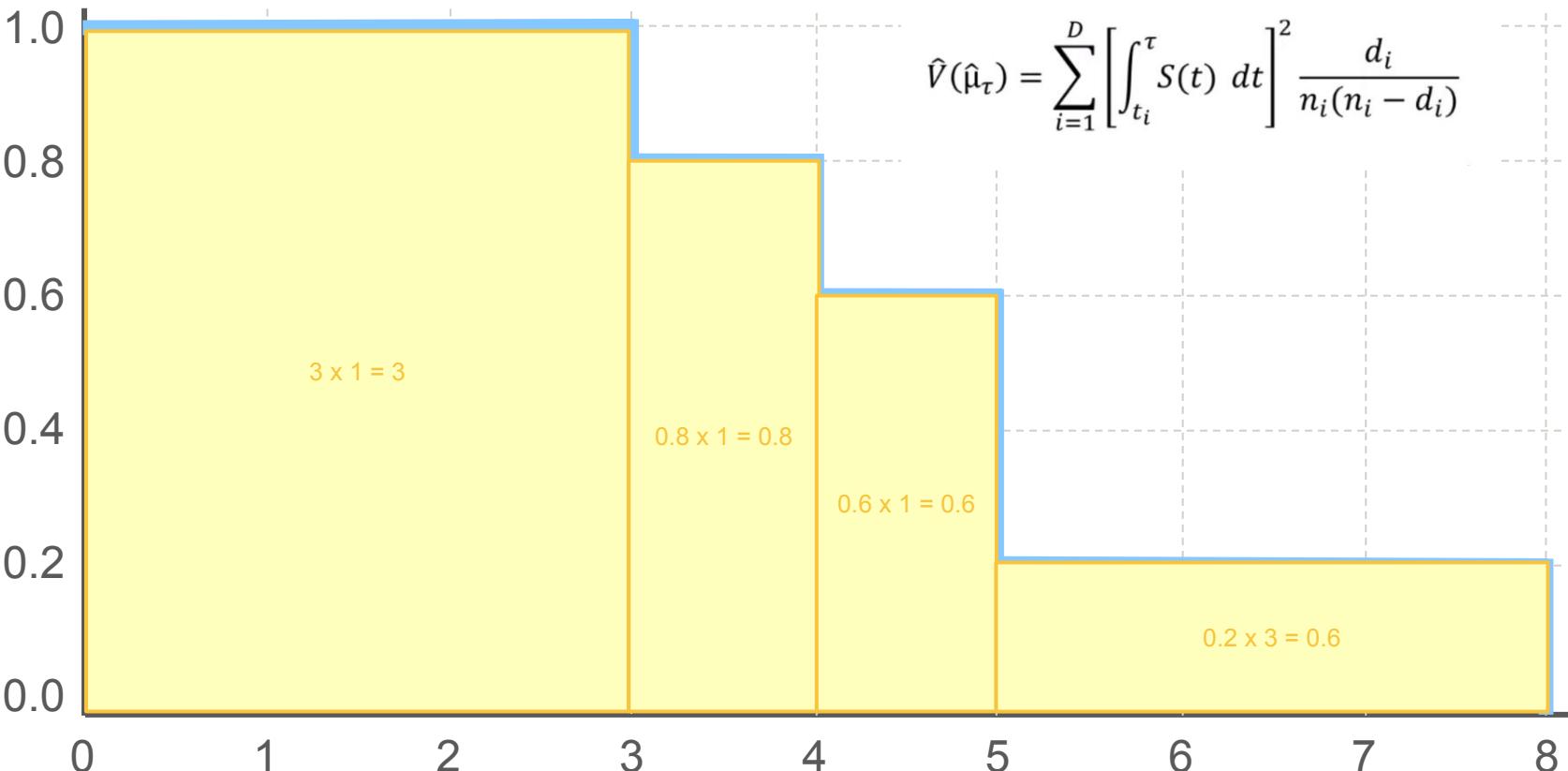
Restricted Mean

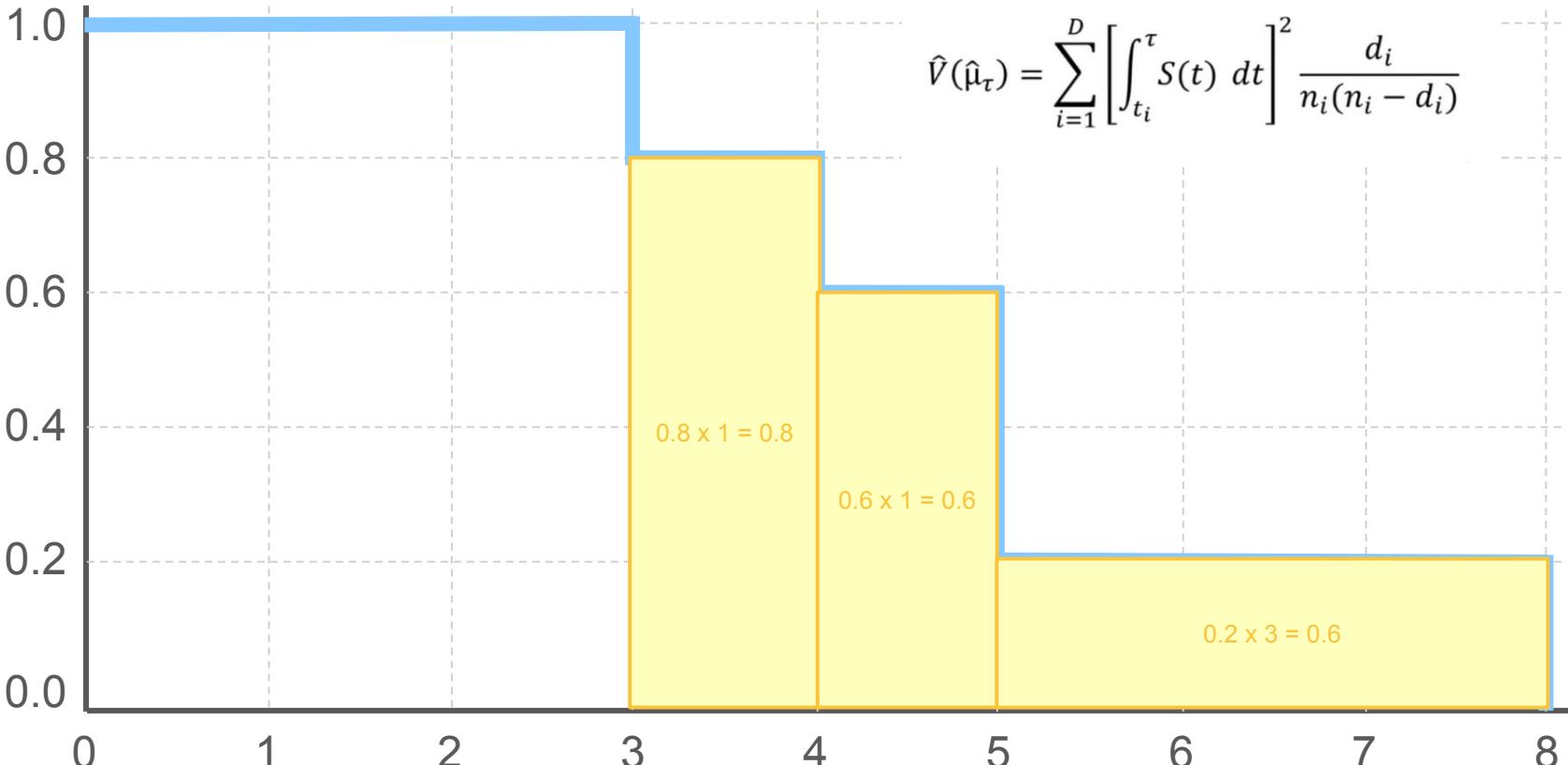
Bootstrap / resampling

$$\hat{V}(\hat{\mu}_\tau) = \sum_{i=1}^D \left[\int_{t_i}^\tau S(t) \, dt \right]^2 \frac{d_i}{n_i(n_i - d_i)}$$

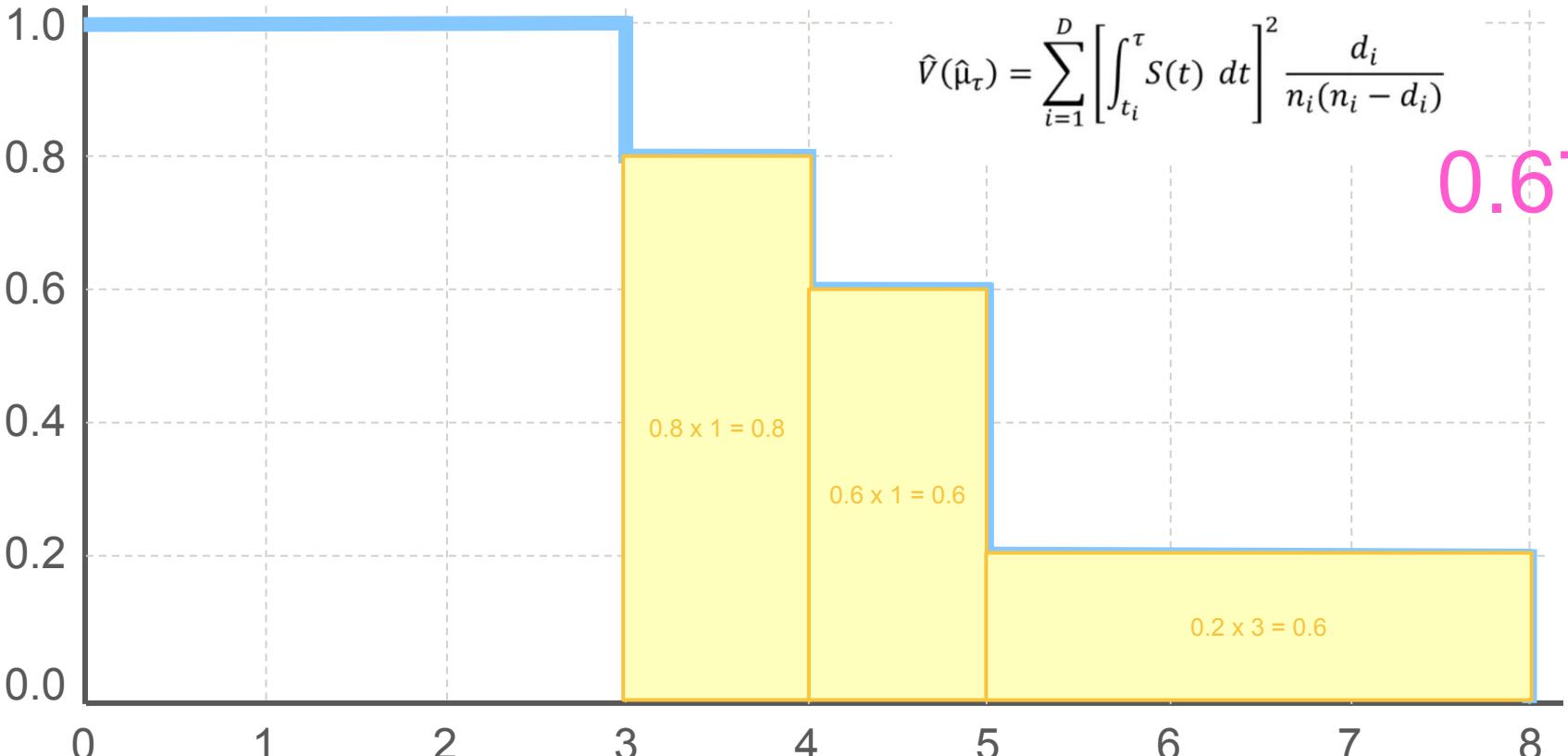
$$CI = \hat{\mu}_\tau \pm Z_{1-\alpha/2} \sqrt{\hat{V}(\hat{\mu}_\tau)}$$

$$\hat{V}(\hat{\mu}_\tau) = \sum_{i=1}^D \left[\int_{t_i}^\tau S(t) \, dt \right]^2 \frac{d_i}{n_i(n_i - d_i)}$$

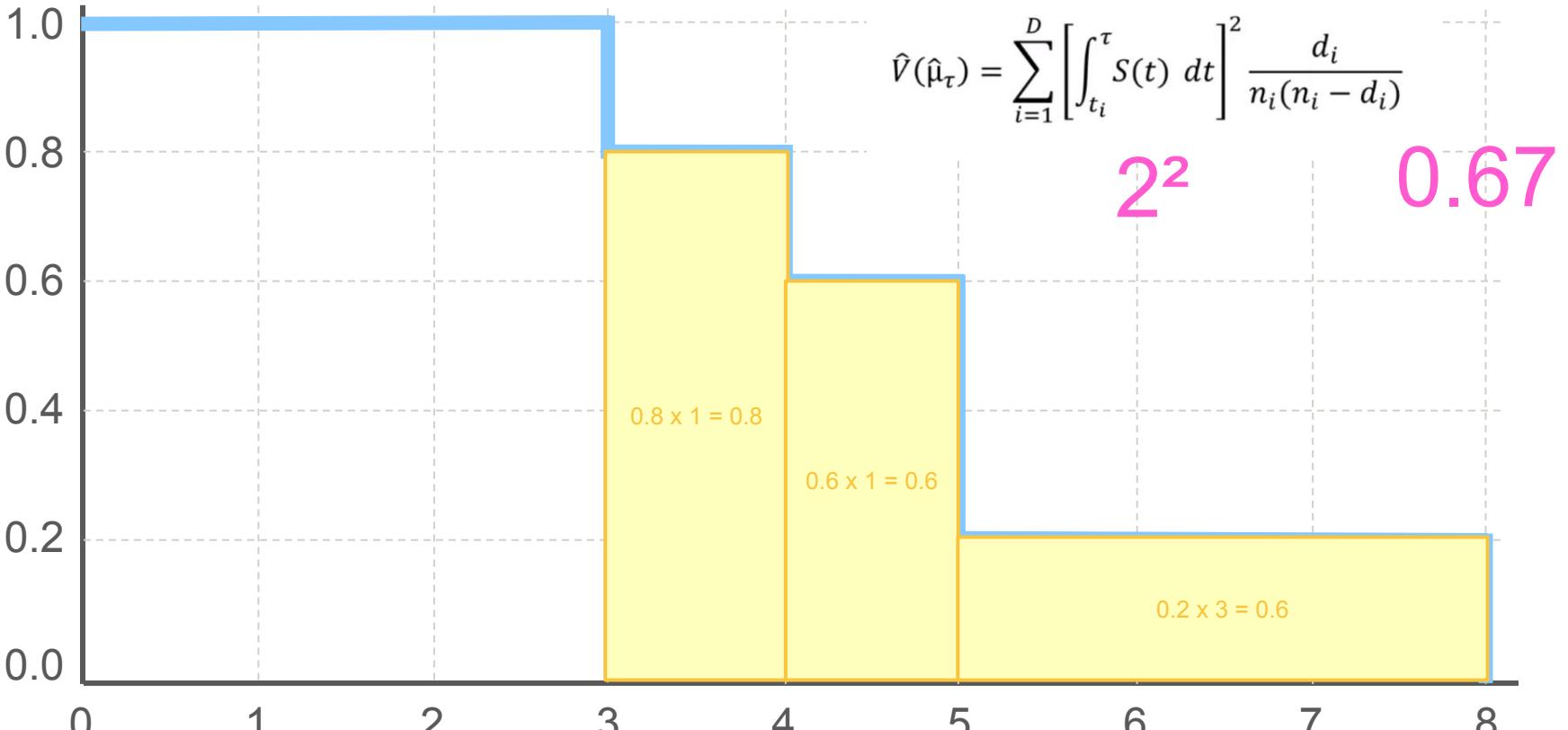




$$0.8 + 0.6 + 0.6 = 2$$

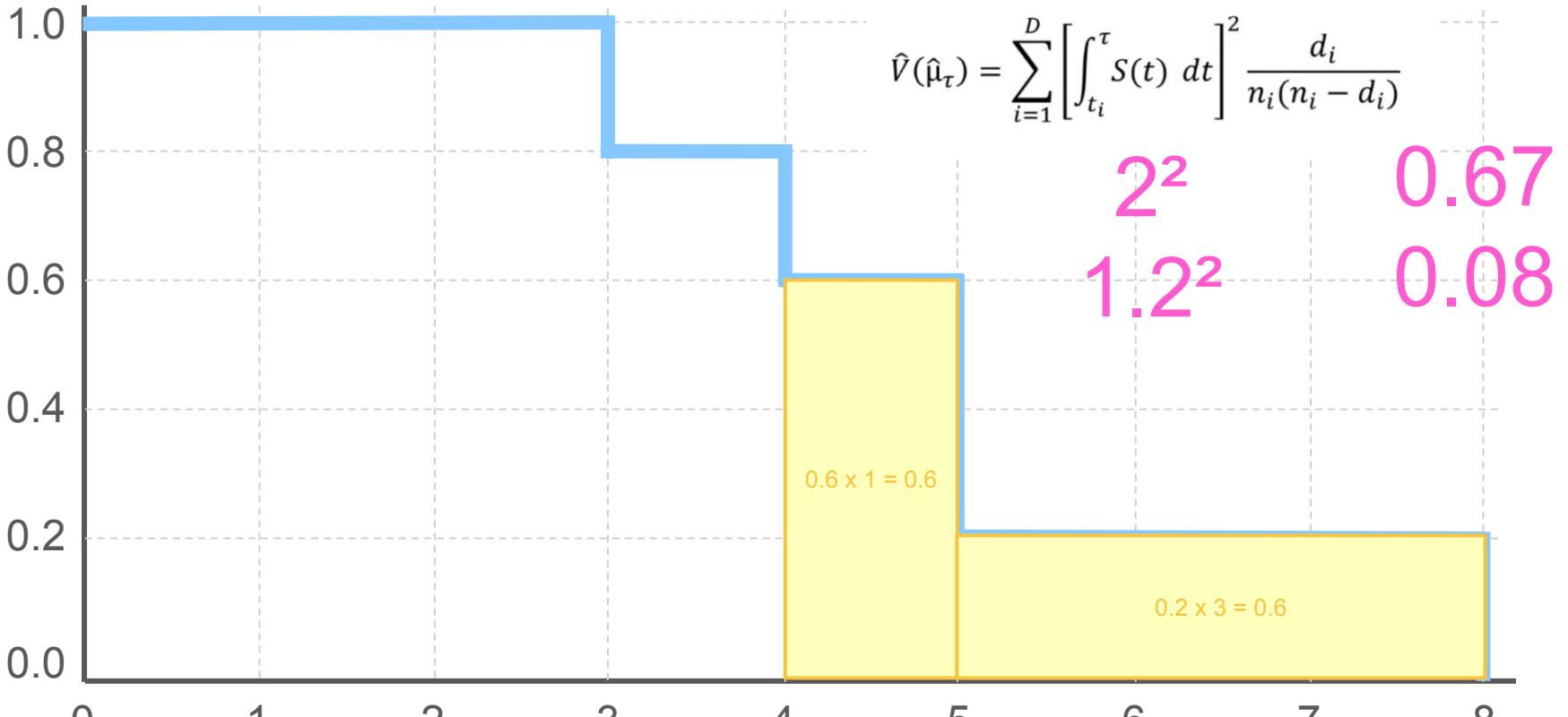


$$0.8 + 0.6 + 0.6 = 2$$

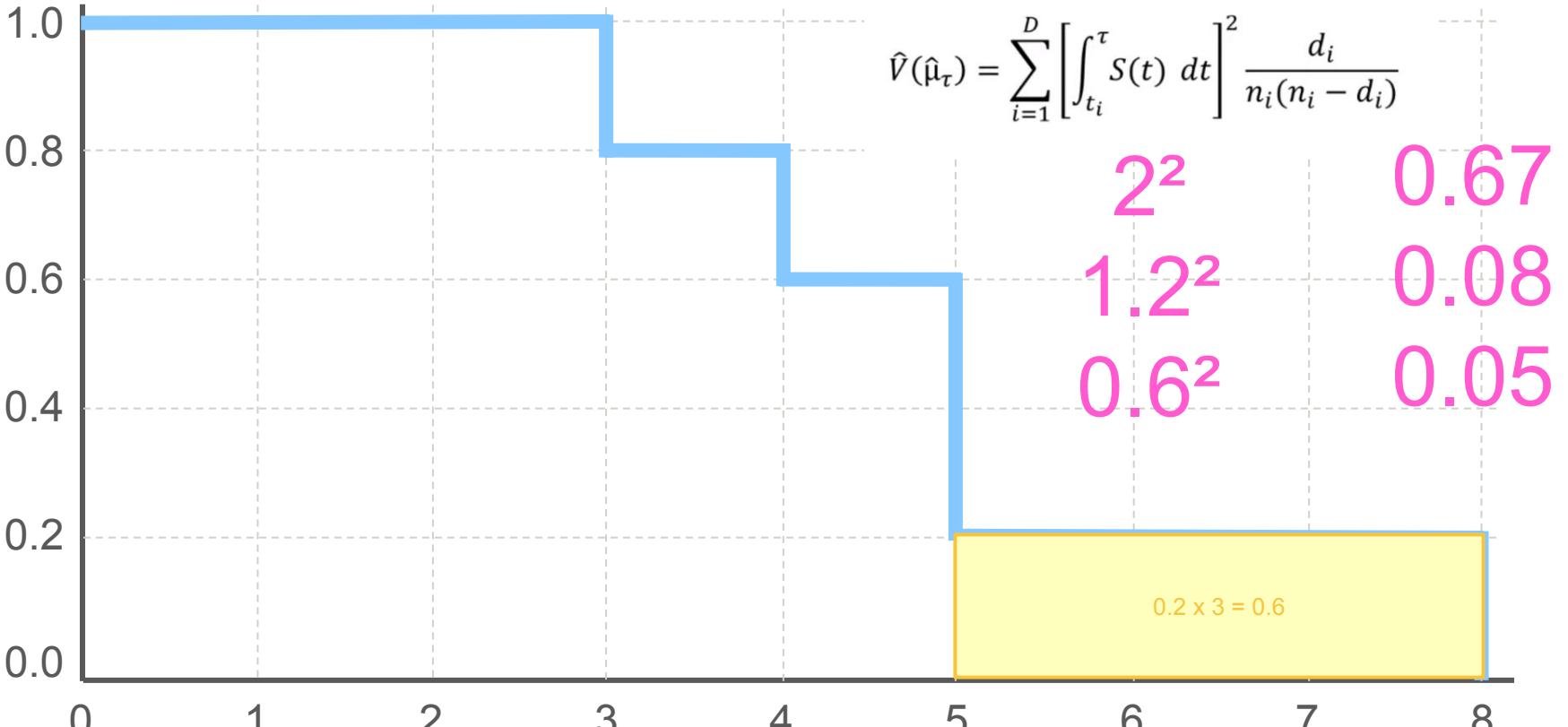


$$0.8 + 0.6 + 0.6 = 2$$

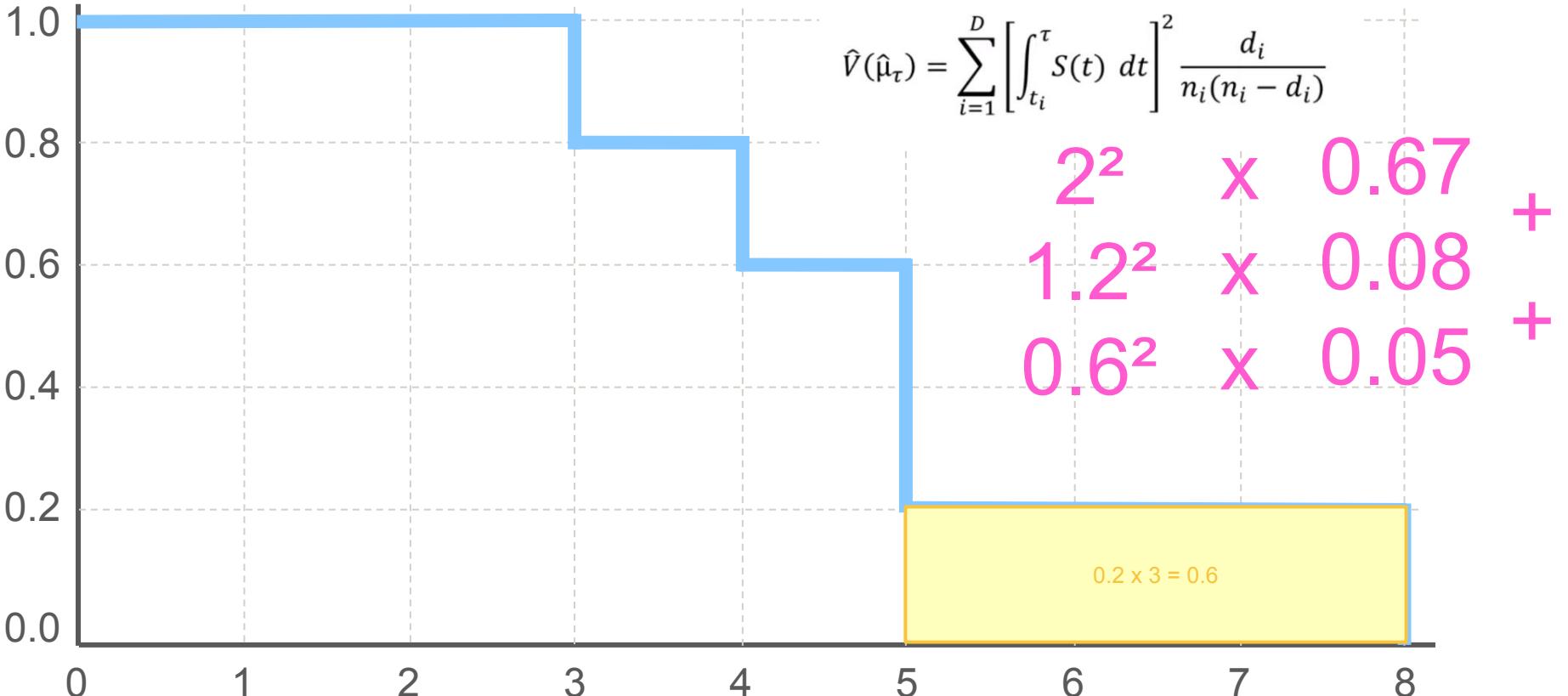
$$\hat{V}(\hat{\mu}_\tau) = \sum_{i=1}^D \left[\int_{t_i}^\tau S(t) \, dt \right]^2 \frac{d_i}{n_i(n_i - d_i)}$$



$$0.6 + 0.6 = 1.2$$



$$0.6 = 0.6$$



$$CI = \hat{\mu}_\tau \pm Z_{1-\alpha/2} \sqrt{0.56}$$

$$CI = 5 \pm Z_{1-\alpha/2} \sqrt{0.56}$$

$$CI = 5 \pm 1.96 \sqrt{0.56}$$

$$3.53 / 6.46 = 5 \pm 1.96 \sqrt{0.56}$$

The confidence interval provides a measure of
the precision of the Kaplan-Meier estimate, not
the probability of survival for an individual

The confidence interval provides a measure of the precision of the Kaplan-Meier estimate, not the probability of survival for an individual

The confidence interval provides a measure of the precision of the Kaplan-Meier estimate, not the probability of survival for an individual

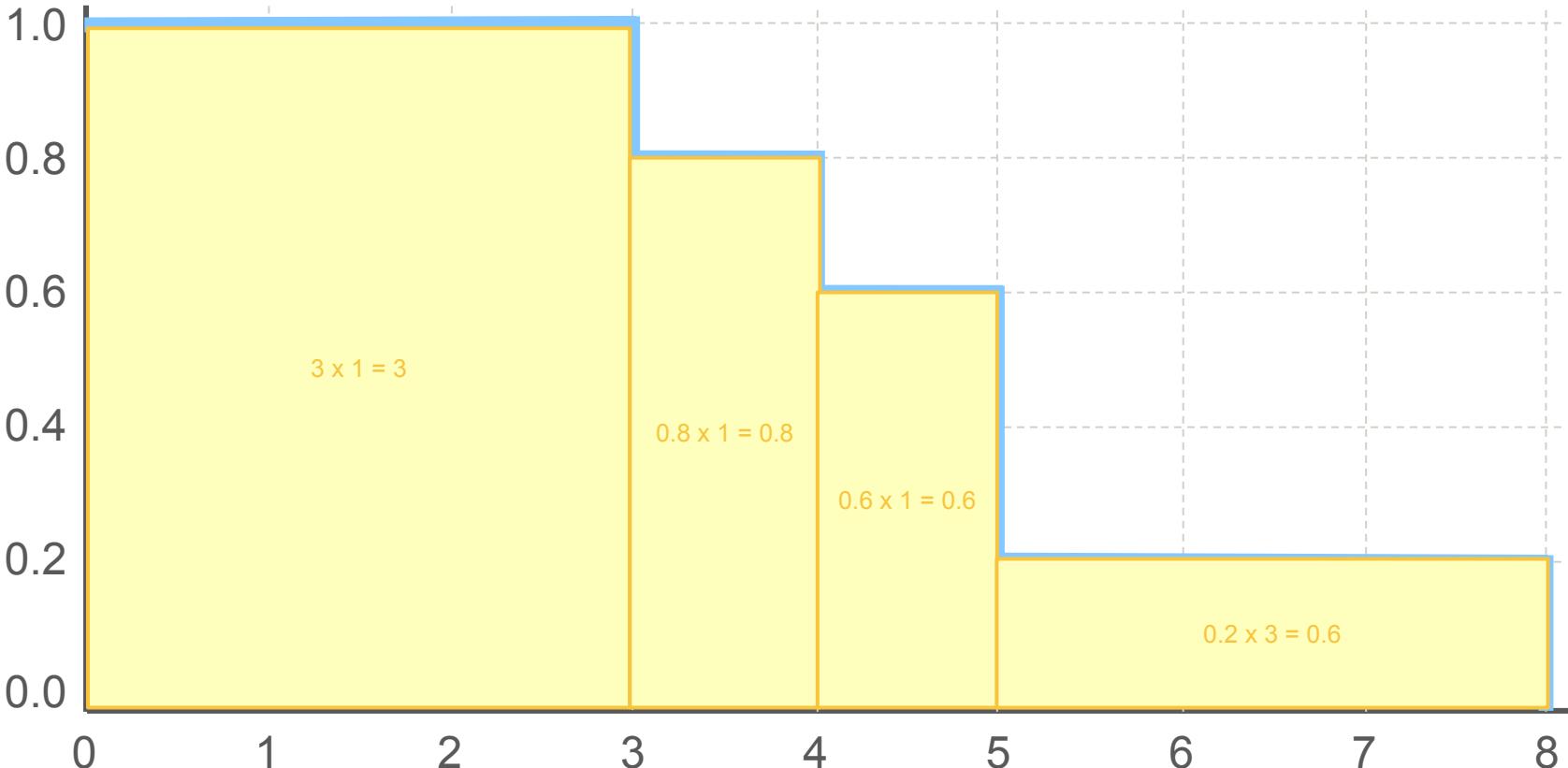


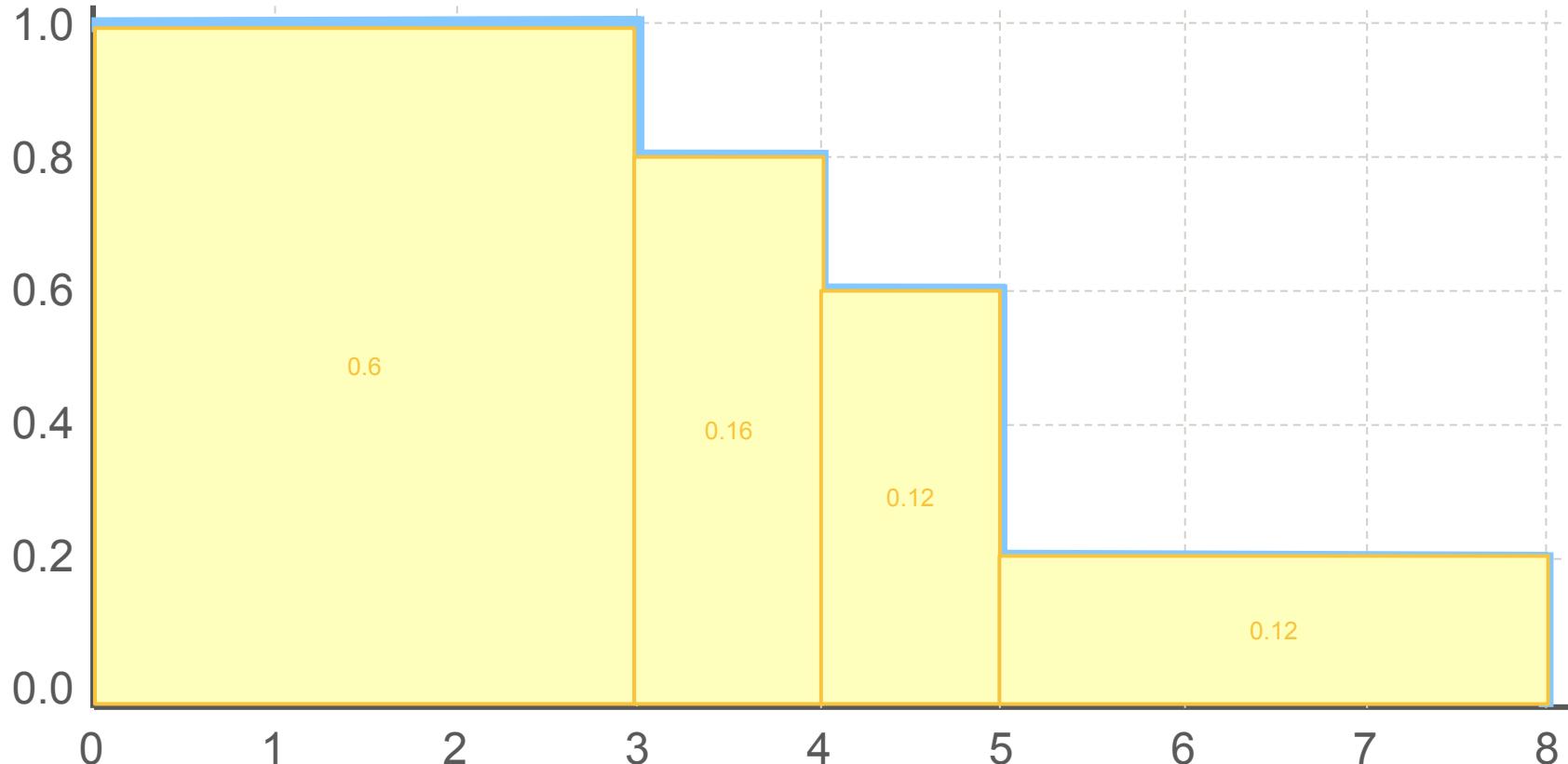
“Oh, the mean probability of survival is 95% between 3.5 and 6.5!”

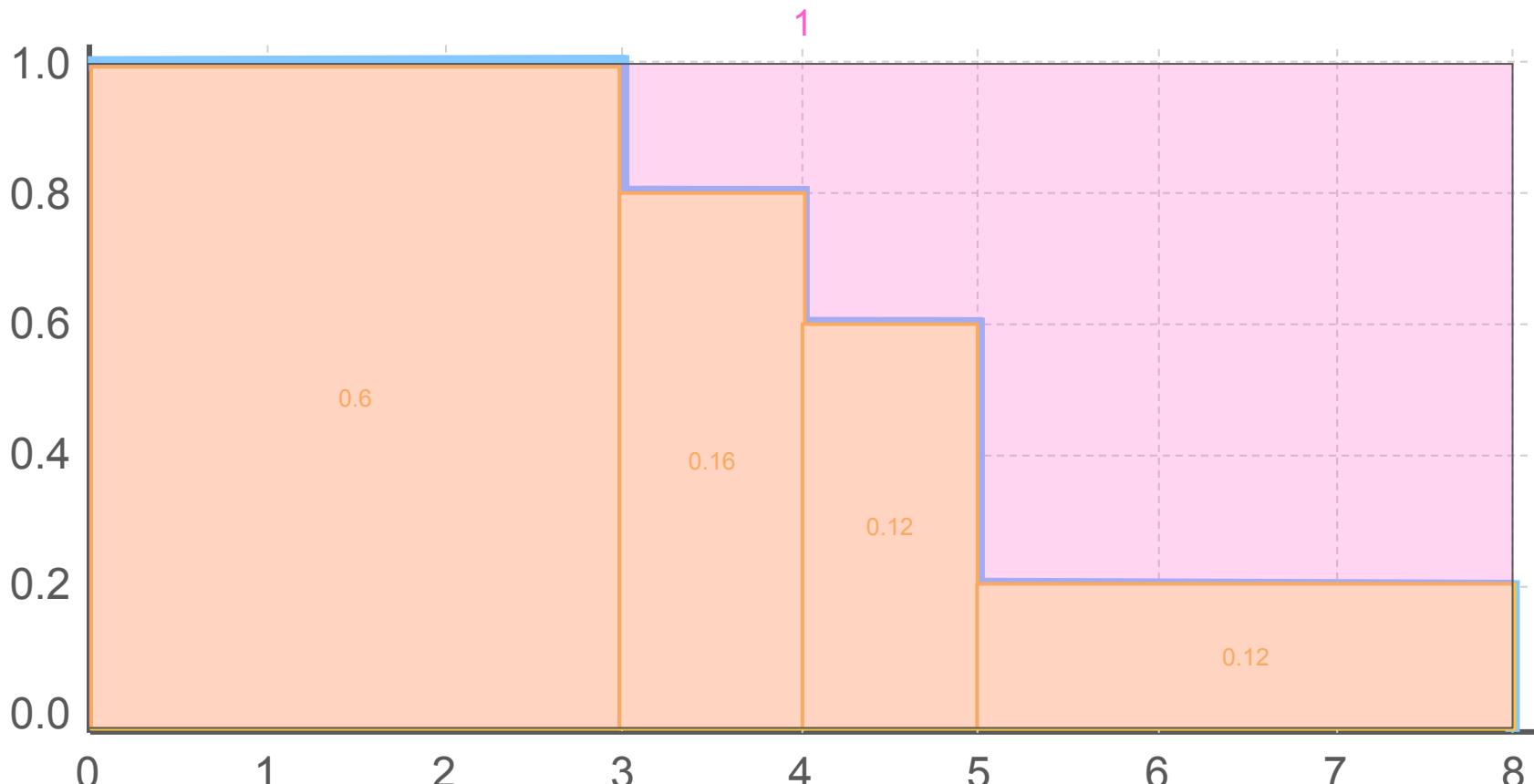
The confidence interval provides a measure of
the precision of the Kaplan-Meier estimate, not
the probability of survival for an individual

*“Oh, the mean probability of survival is
95% between 3.5 and 6.5!”*



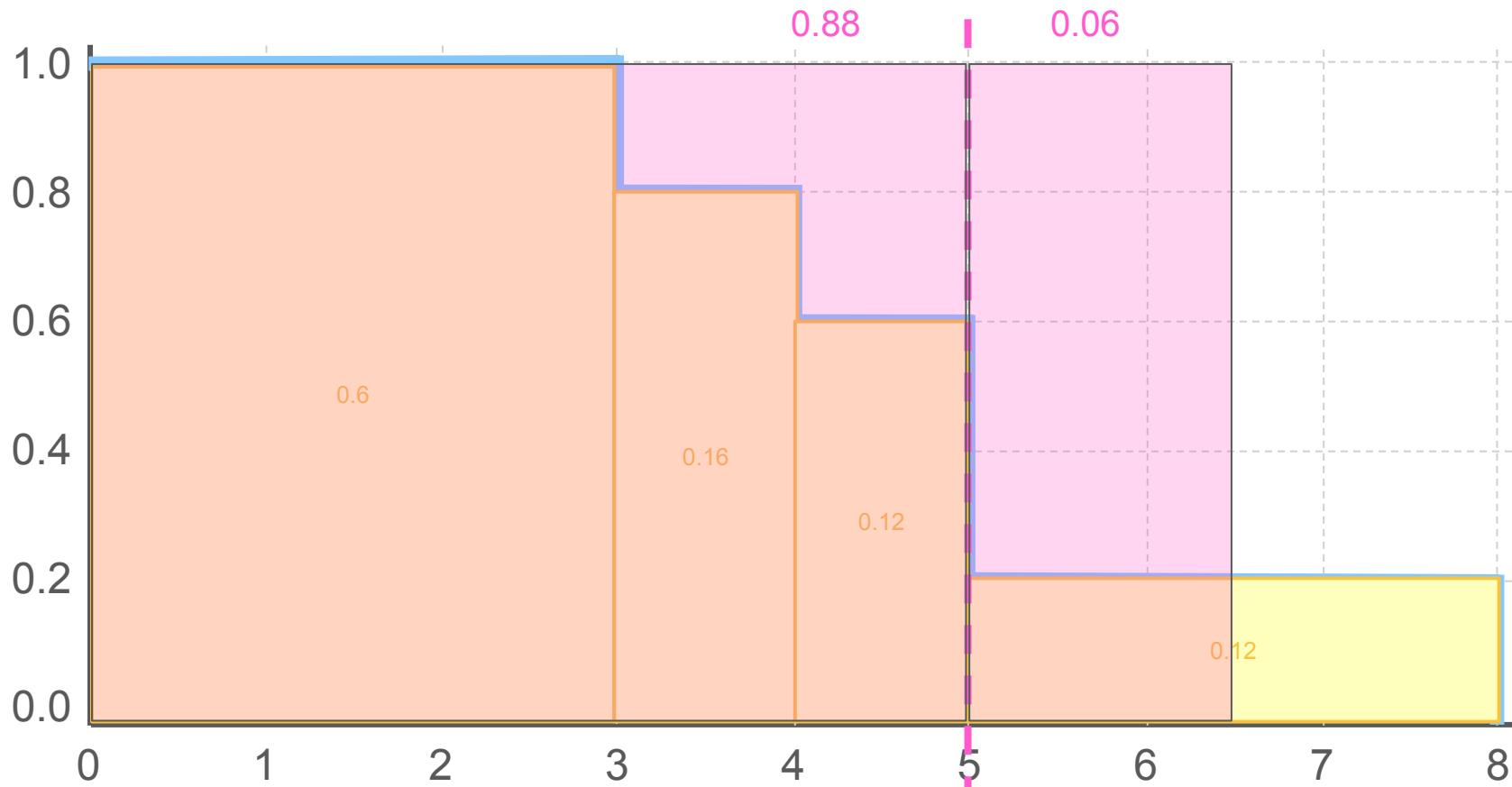






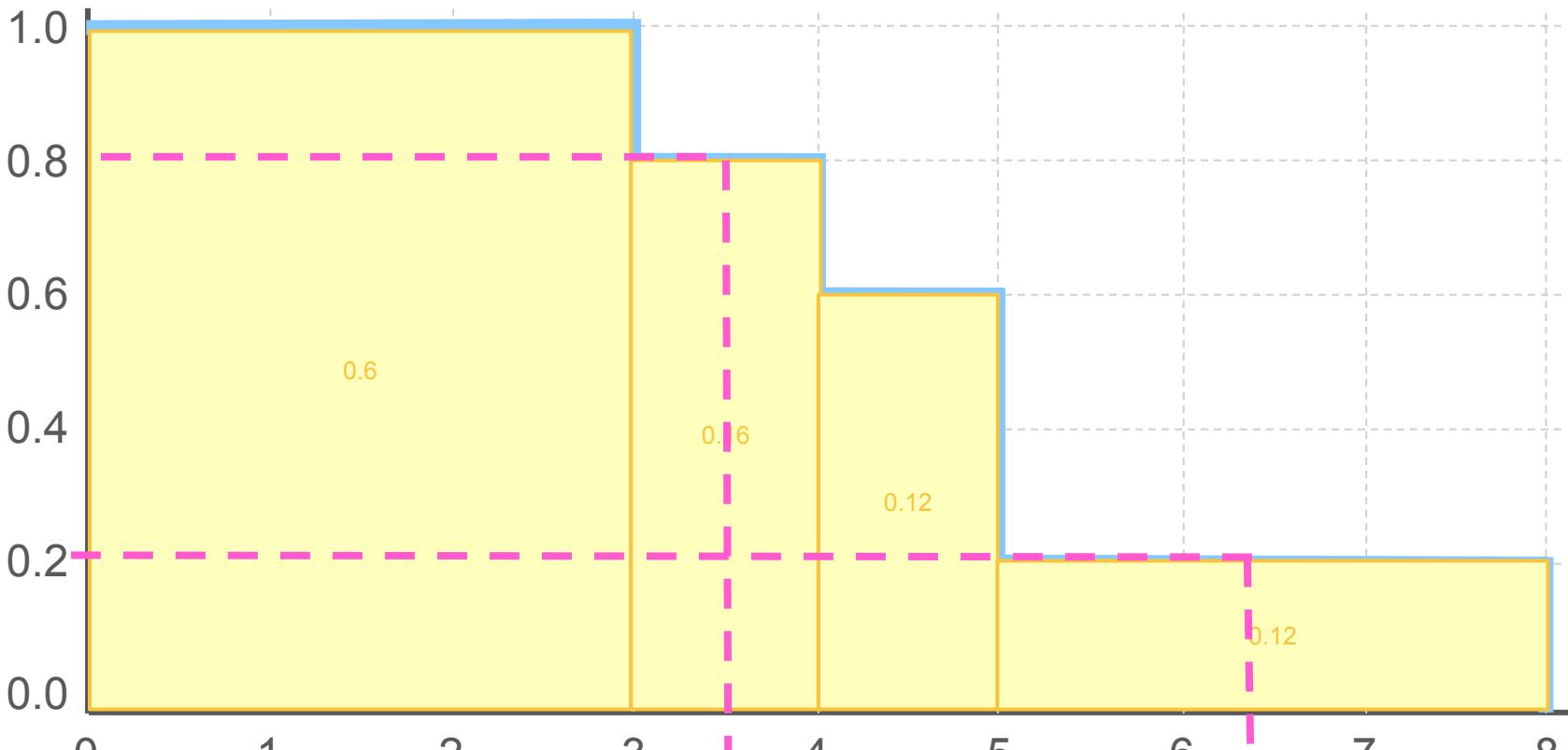
$$\int f(x) = 1$$

100% to last 0 to 8 years



$$\int f(x) = 0.94$$

94% to last 0 to 6.5 years



$$S(3.5) = 0.8$$

80% survival
at 3.5 years

$$S(6.5) = 0.2$$

20% survival
at 6.5 years

2 - SURVIVAL CURVES



Survival curves

Kaplan-Meier



Survival curves

Non-parametric

Kaplan-Meier

Parametric

Exponential

Weibull

Log-Normal



Survival curves

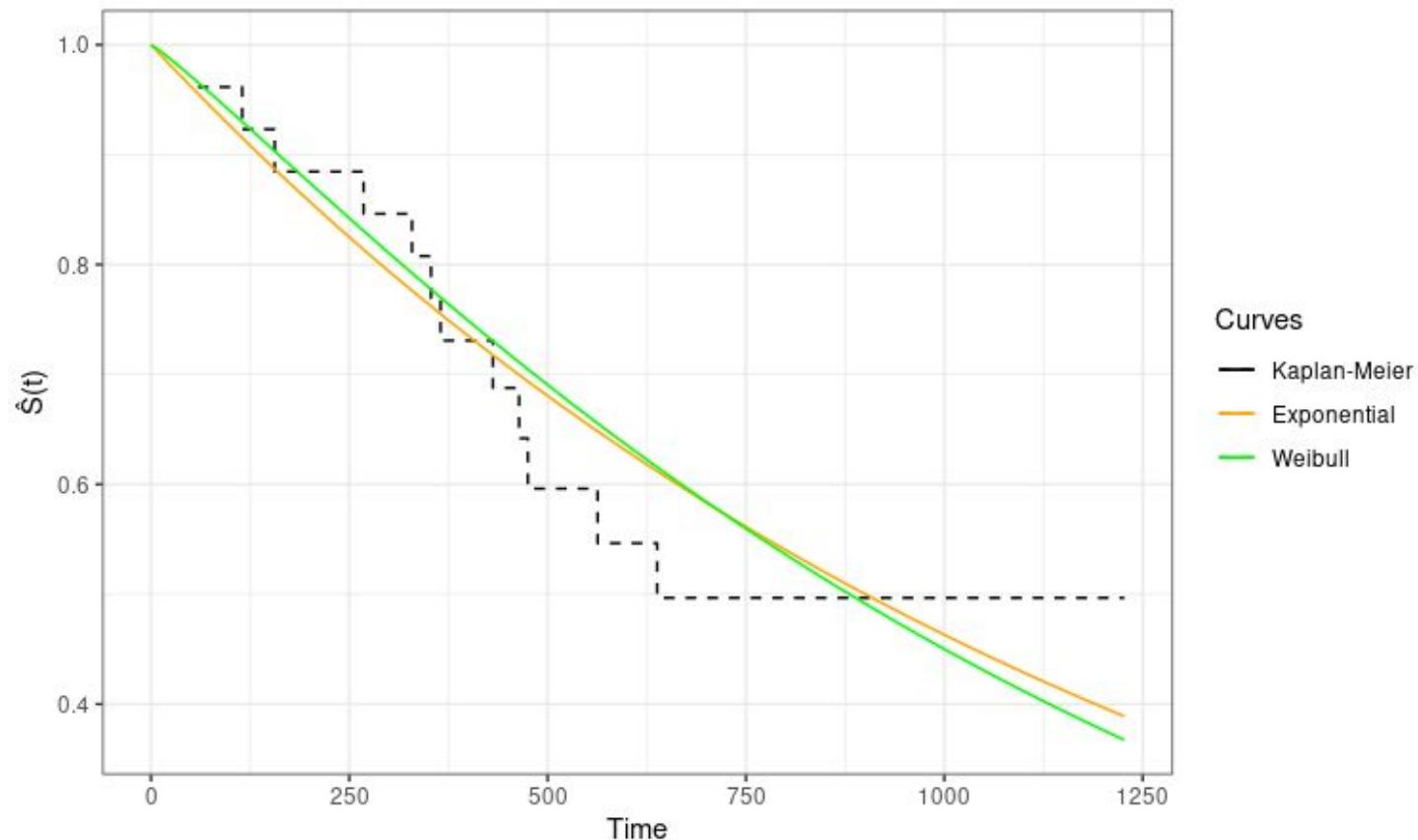
Non-parametric

Doesn't assume distribution

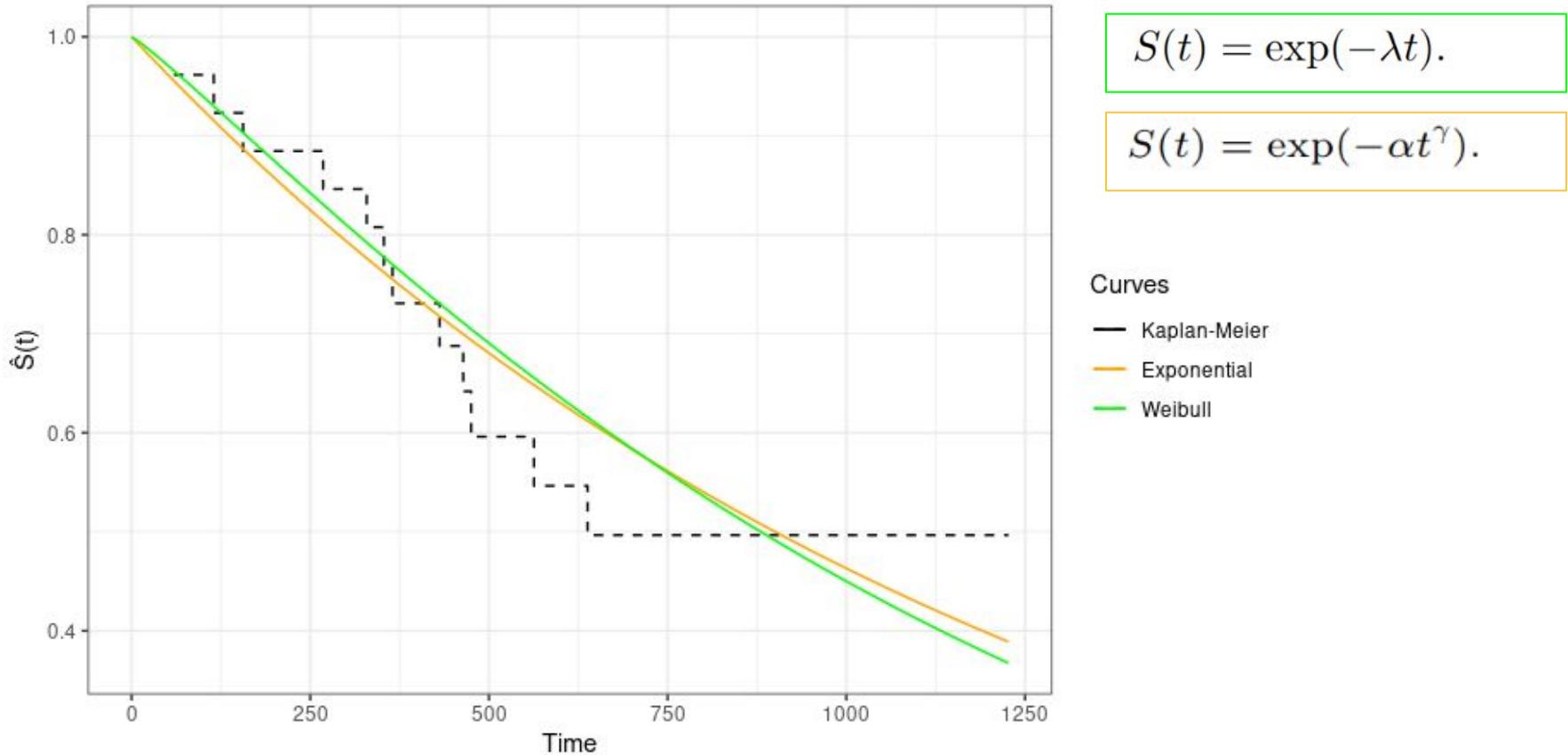
Parametric

Data must follow distribution
and you need to figure out the
parameters for it

KM - Exponential - Weibull comparison



KM - Exponential - Weibull comparison



Maximum likelihood estimation (MLE)

asymptotically unbiased, consistent, and efficient

Method of Moments

not efficient and not handling censored data as naturally

Bayesian Estimation

useful with knowledge, complex models, or small sample size (see Markov Chain Monte Carlo)

Penalized Likelihood

add a penalty to prevent overfitting (see Lasso and Ridge)

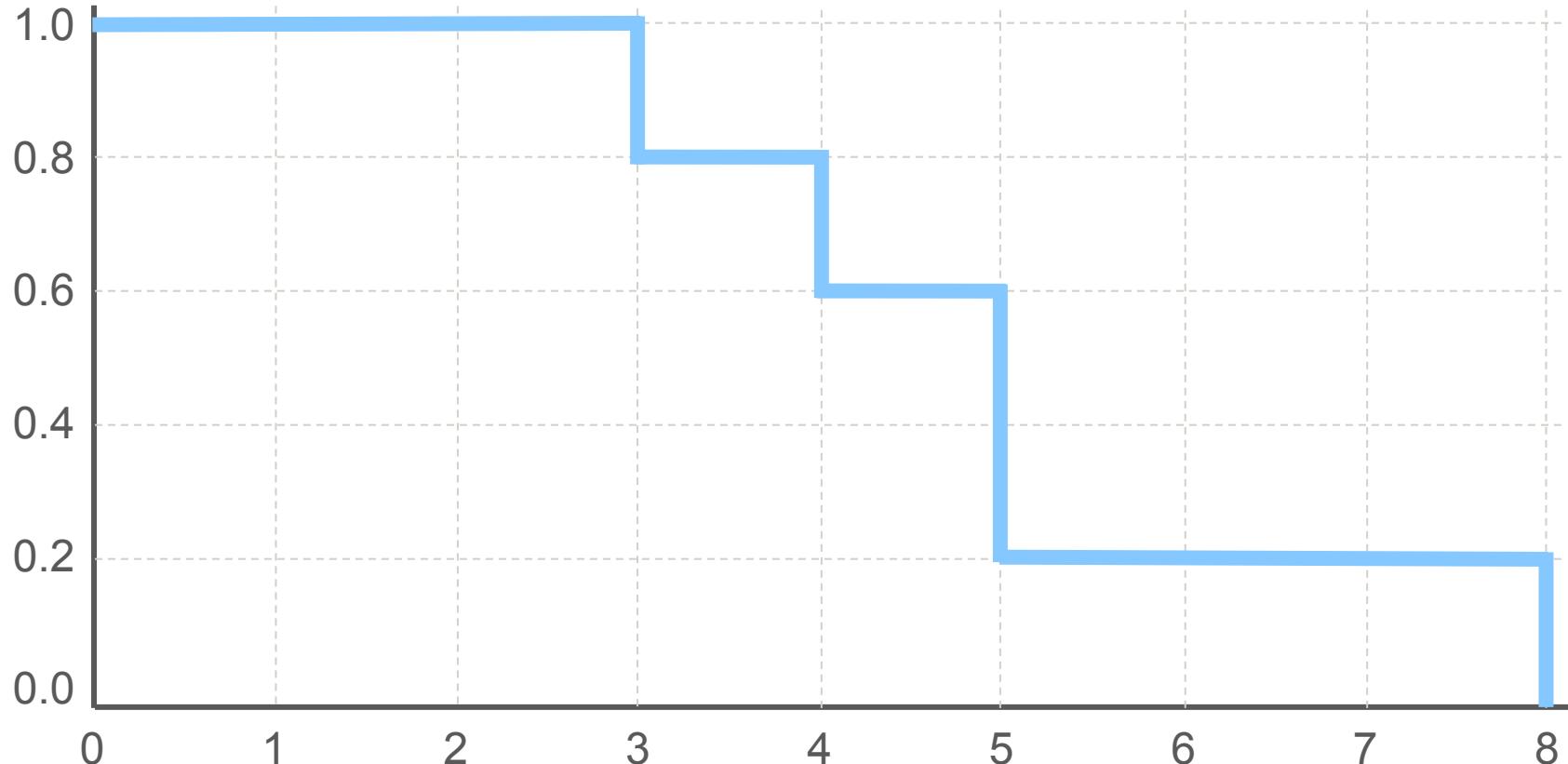
Least Squares

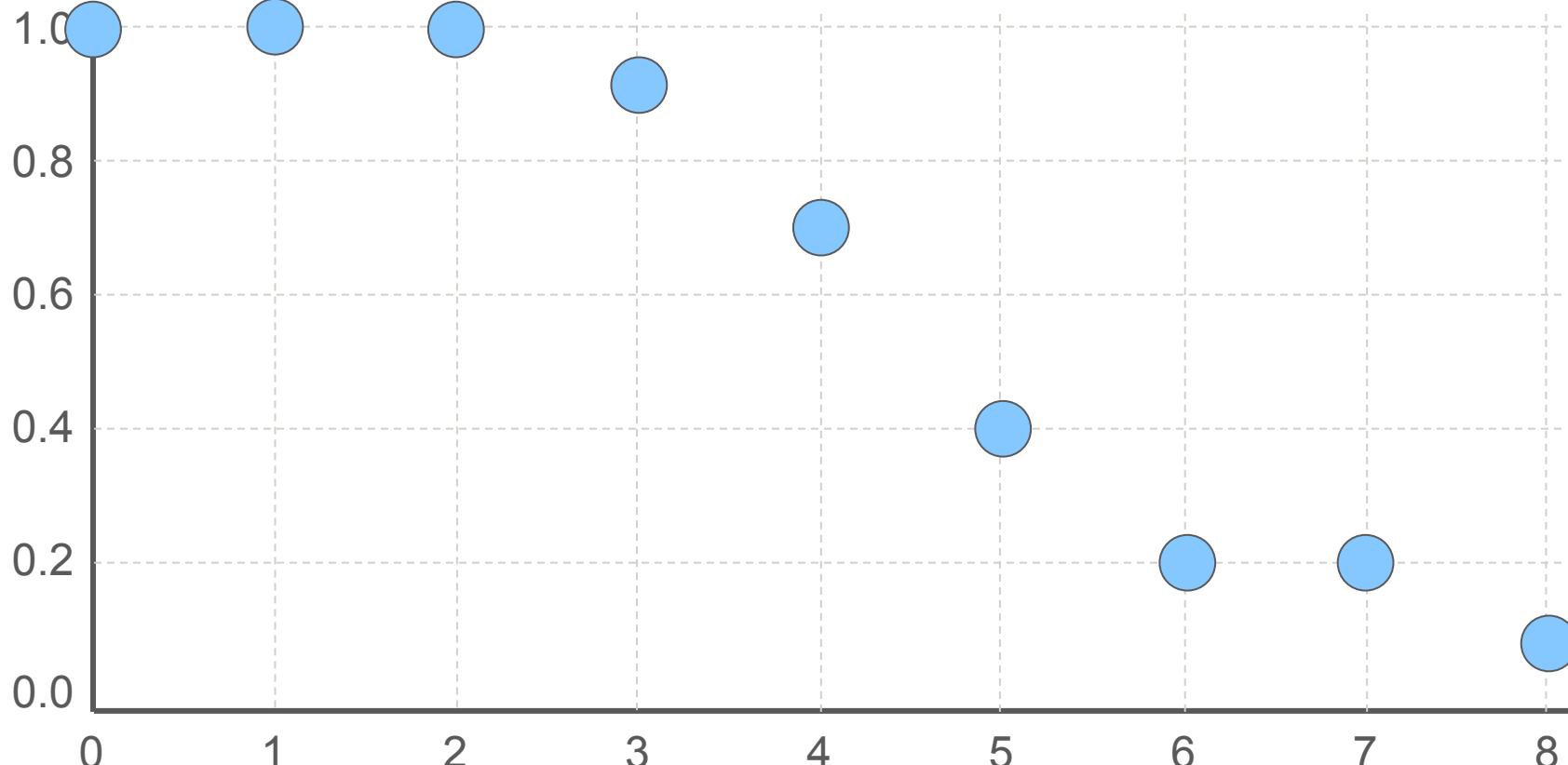
when fitting linear regression to survival data. Bad censoring.

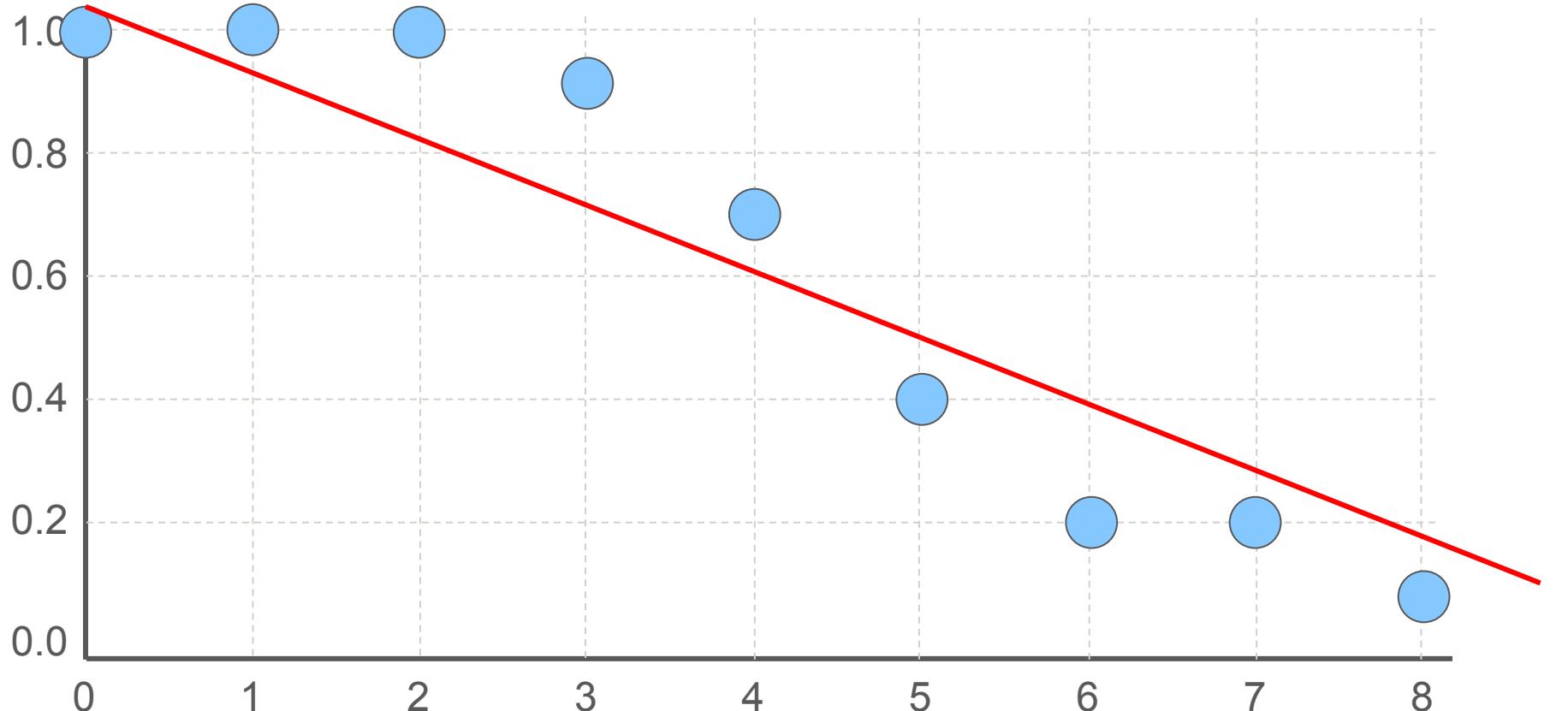
Rank-Based Estimation

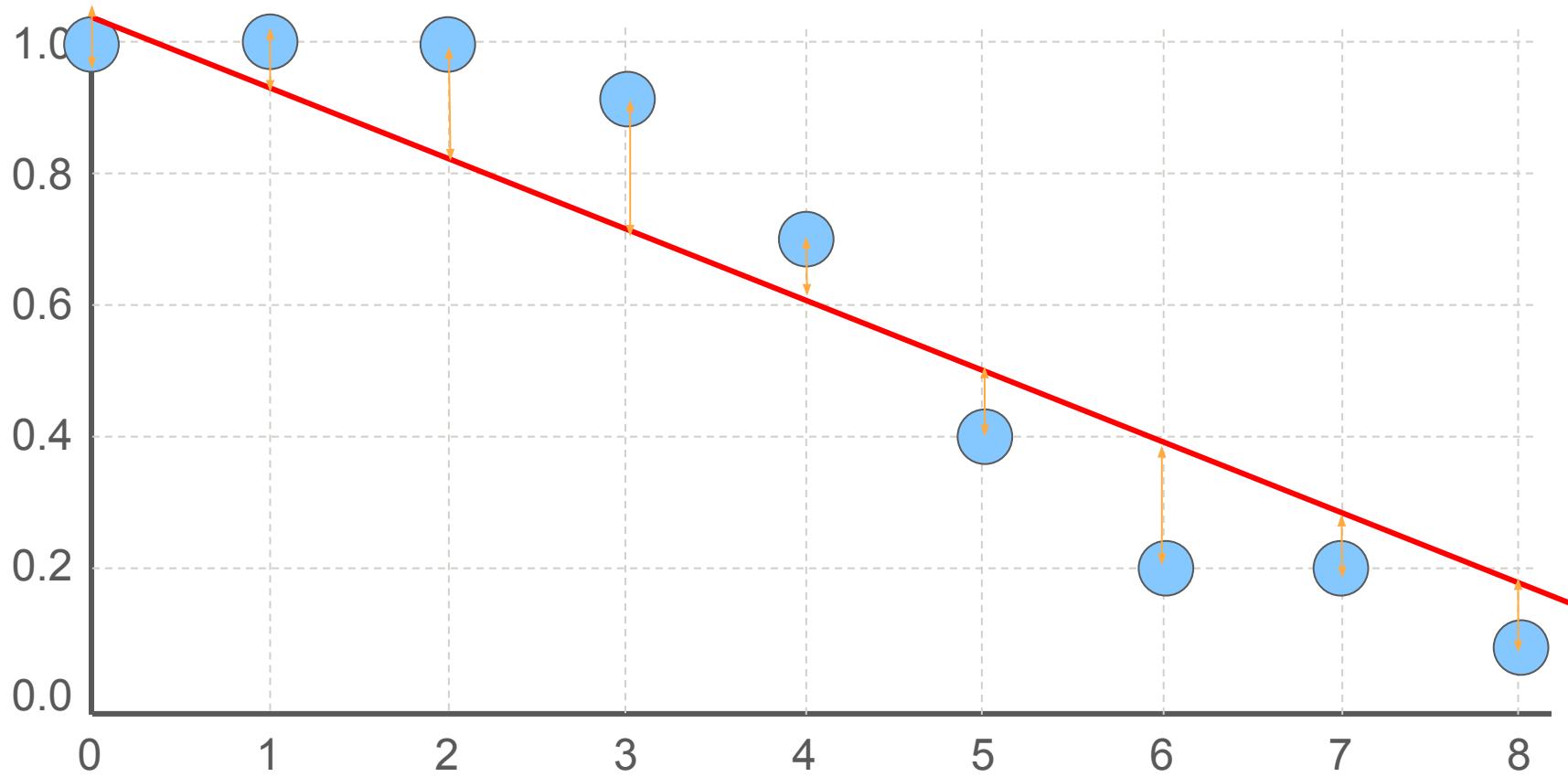
Cox's proportional hazards model (later)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900	901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919	920	921	922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	939	940	941	942	943	944	945	946	947	948	949	950	951	952	953	954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978	979	980	981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	1000



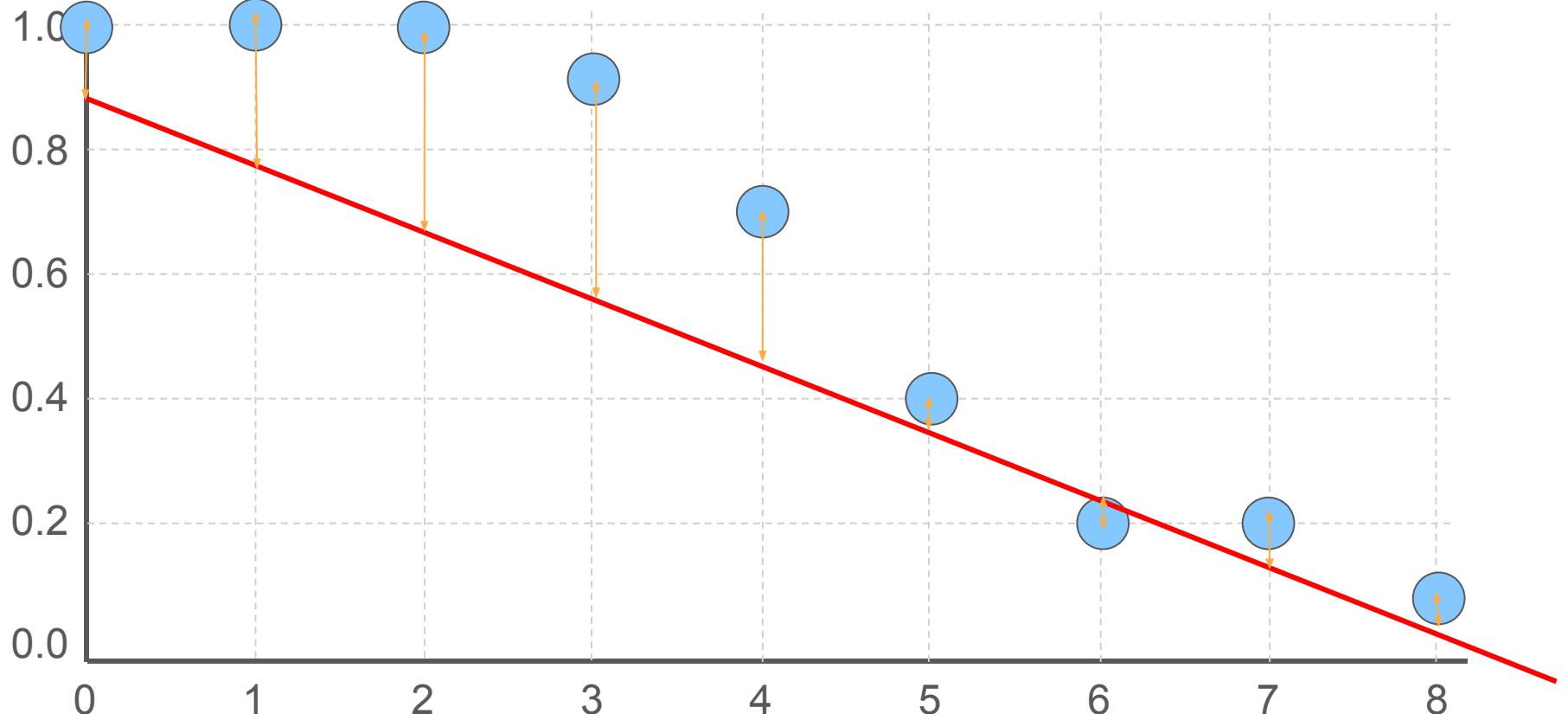






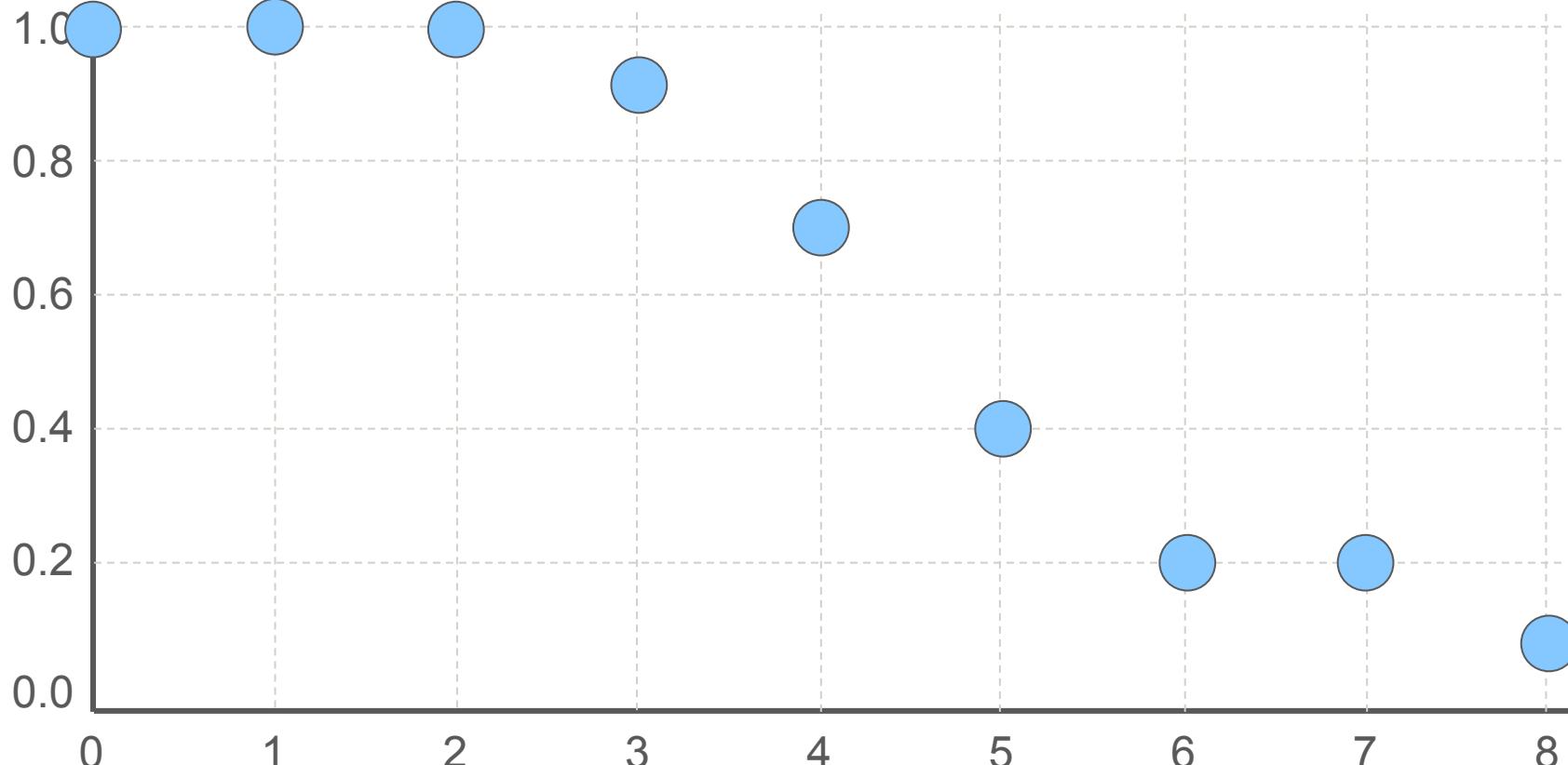
Minimum Least (Square)

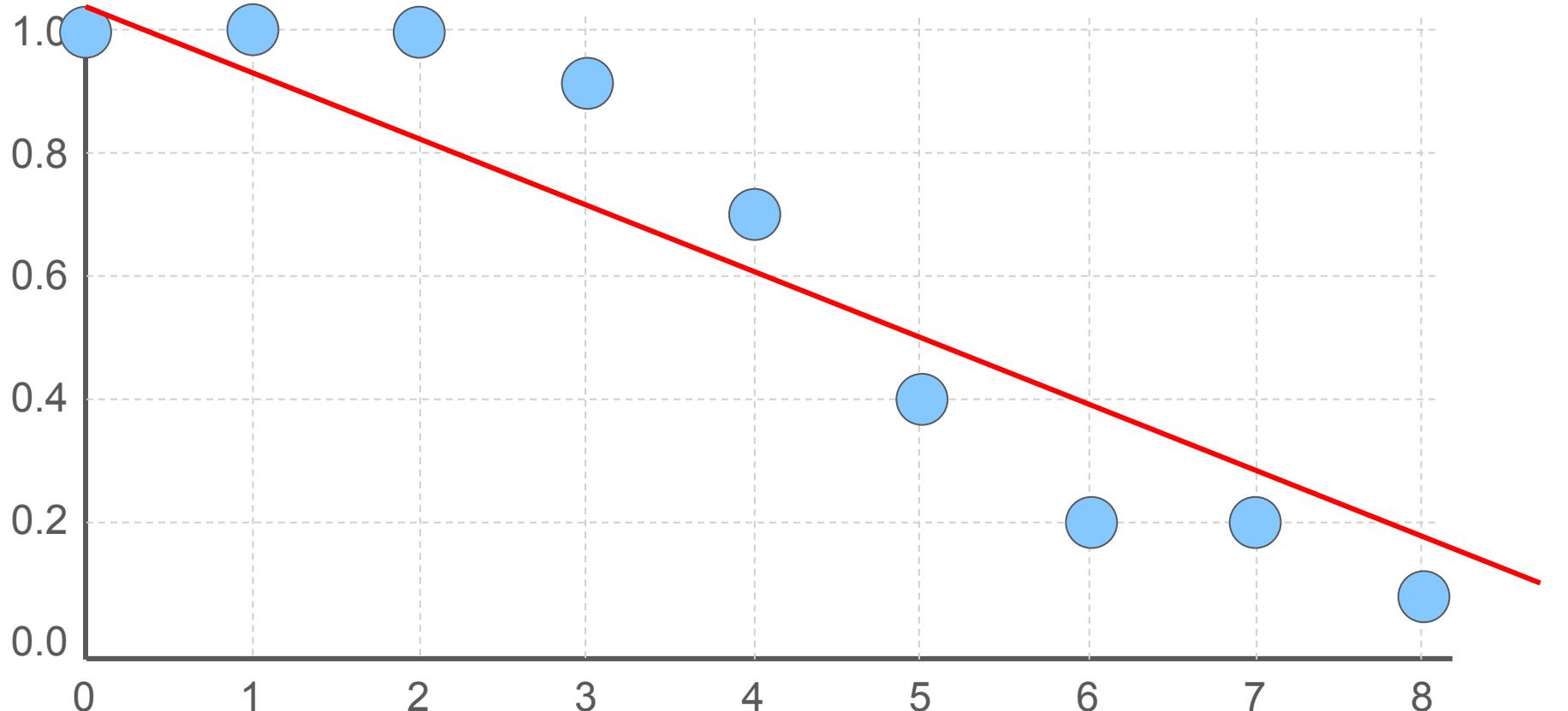


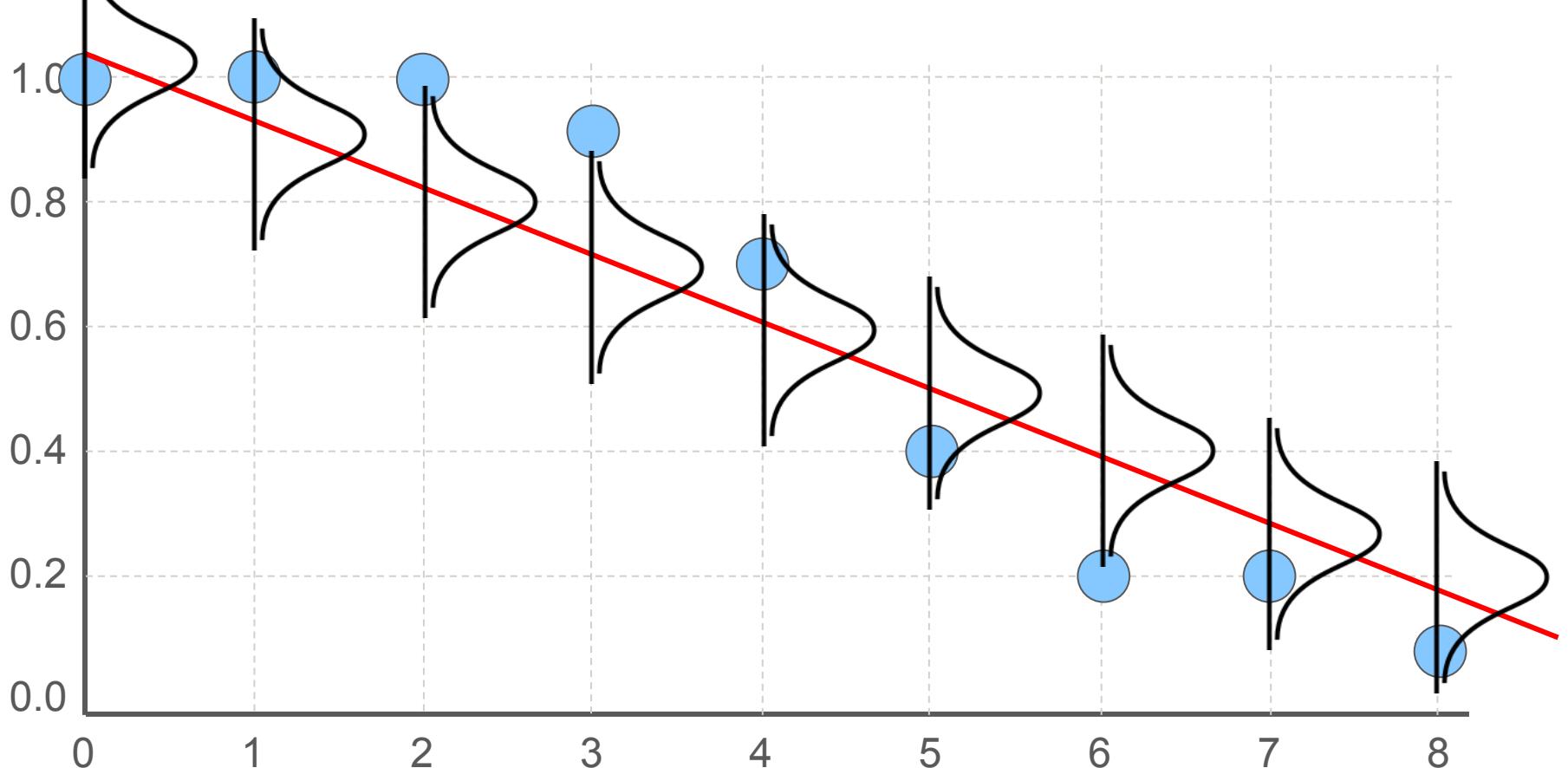


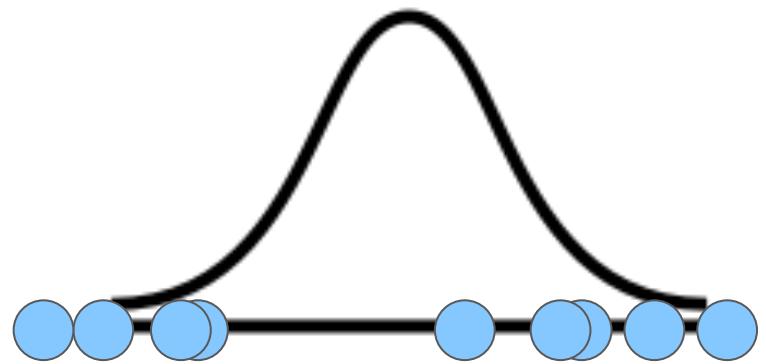
Minimum Least (Square)

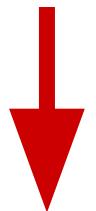
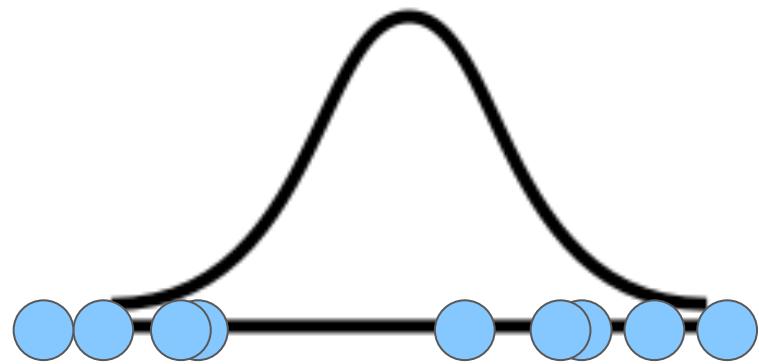




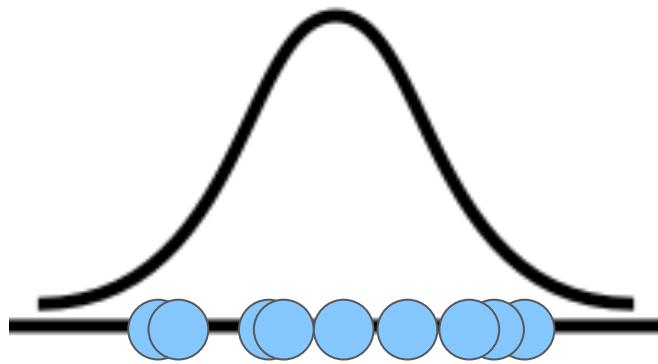


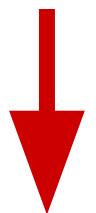
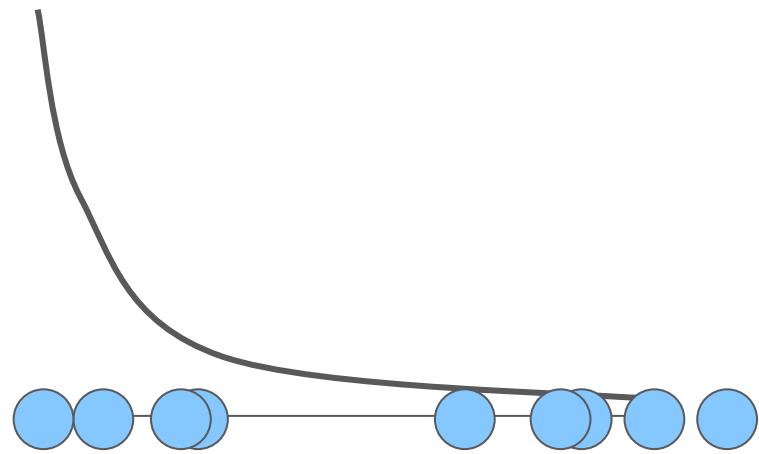




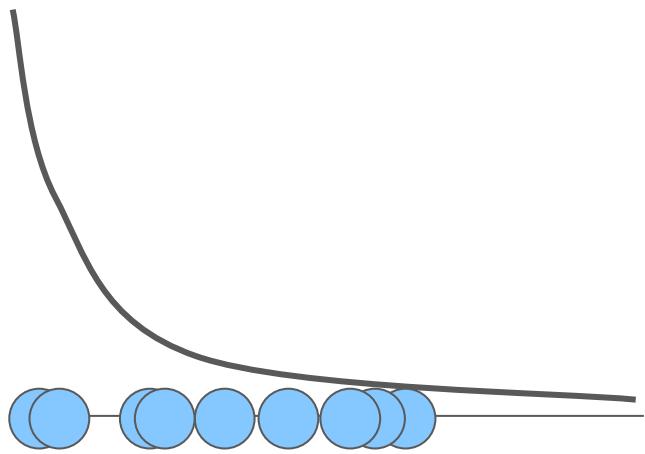


Π Probability

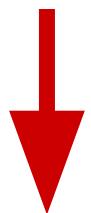
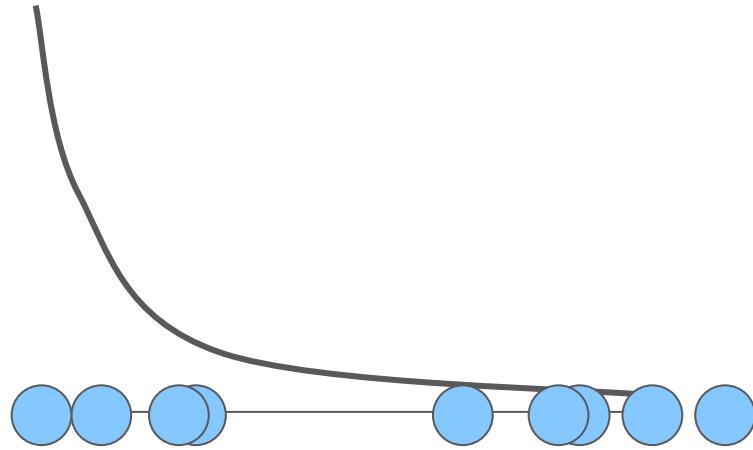




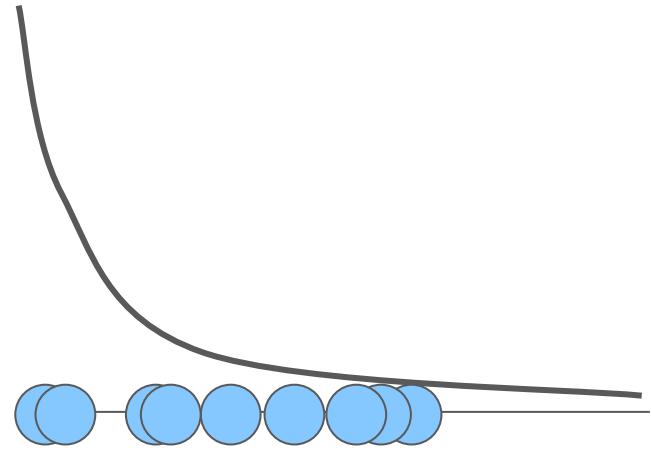
Π Probability



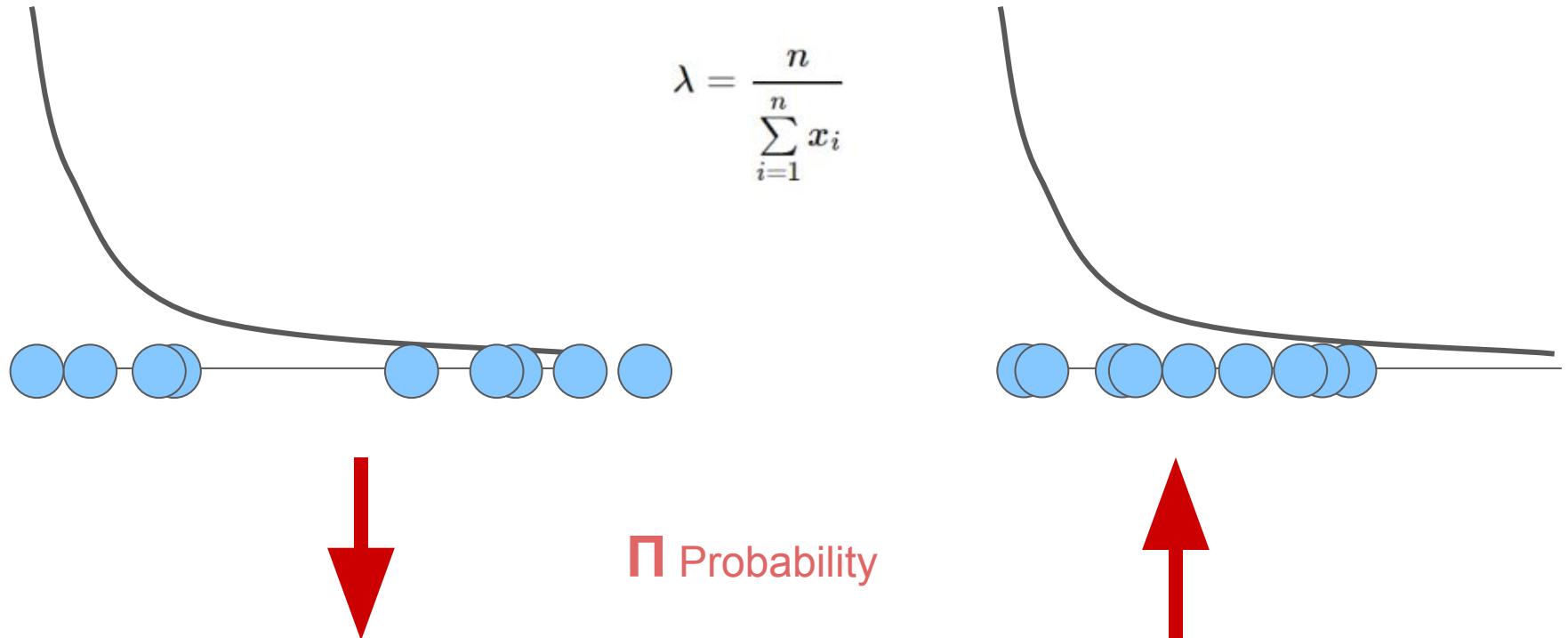
$$\mathcal{L}(\lambda, x_1, \dots, x_n) = \prod_{i=1}^n f(x_i, \lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^n x_i}$$



Π Probability



$$\mathcal{L}(\lambda, x_1, \dots, x_n) = \prod_{i=1}^n f(x_i, \lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^n x_i}$$



3 - TIME TO EVENT



UiT The Arctic University of Norway

SURVIVAL ANALYSIS

An introduction to basic concepts

Rafael Nozal Cañadas

HDL Seminar 2024.01.29





UiT The Arctic University of Norway

Time-to-Event ~~SURVIVAL ANALYSIS~~ An introduction to basic concepts

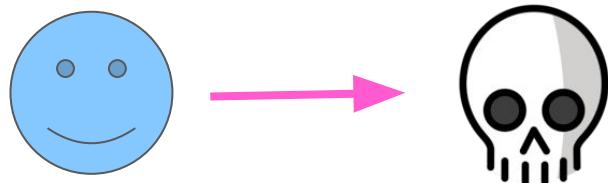
Rafael Nozal Cañas

HDL Seminar 2024.01.29

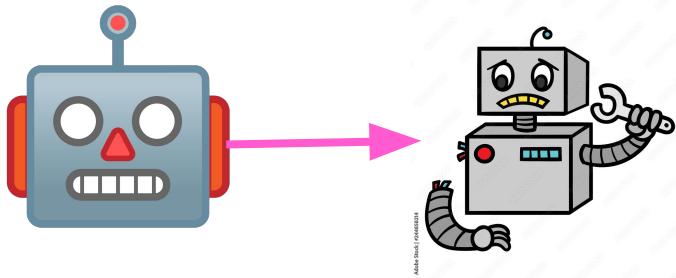




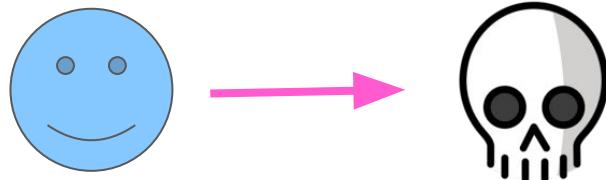
Person dies after some time
(Overall Survival)



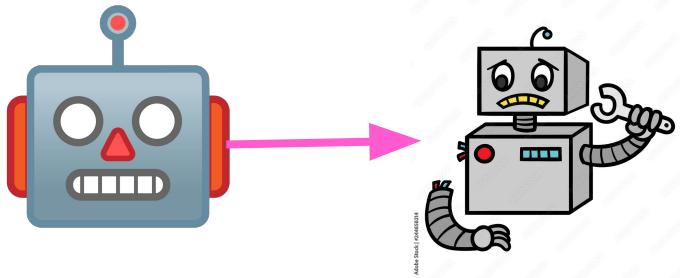
Person dies after some time



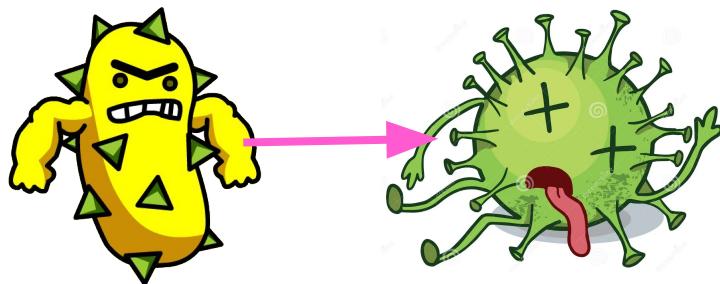
Machine dies after some time
(Reliability Analysis)



Person dies after some time



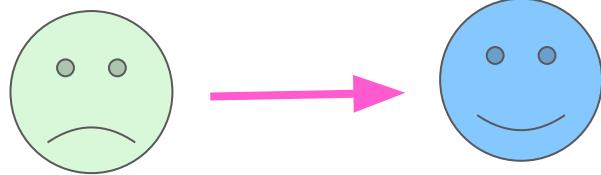
Machine dies after some time



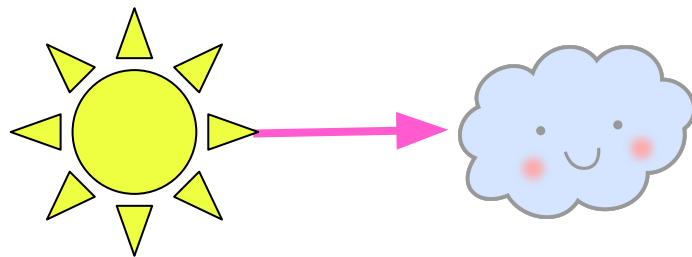
Bacteria dies after some time
(Time to Treatment Failure)



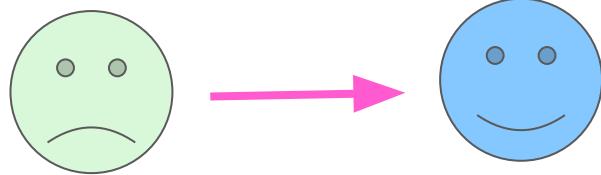
Infected person become
healthy after some time



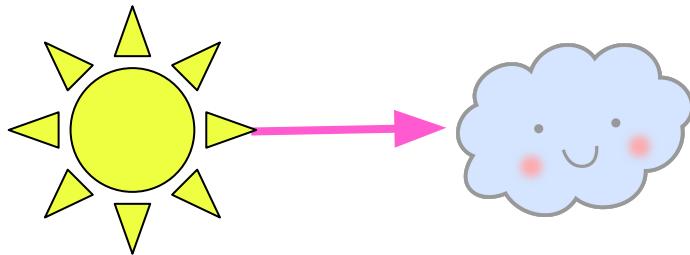
Infected person become
healthy after some time



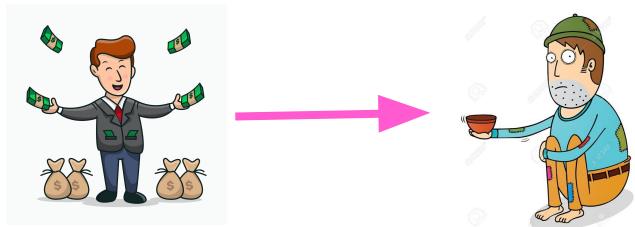
Cloudy after some time
(Event-Free Survival)



Infected person become healthy after some time



Cloudy after some time



Default money after some time
(Time until default)

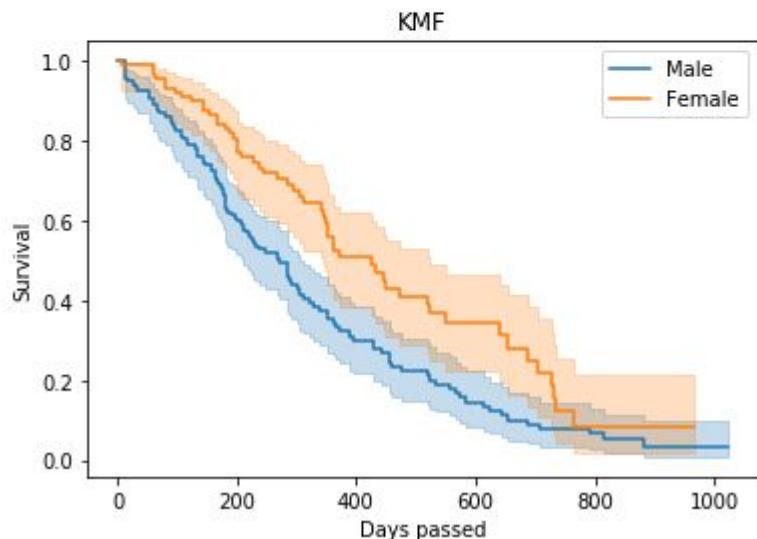


Increase risk = **Bad!**



Increase risk = **Good!**

```
# Plot the survival_function data :  
  
kmf_m.plot()  
kmf_f.plot()  
  
plt.xlabel("Days passed")  
plt.ylabel("Survival")  
plt.title("KMF")  
  
Text(0.5, 1.0, 'KMF')
```

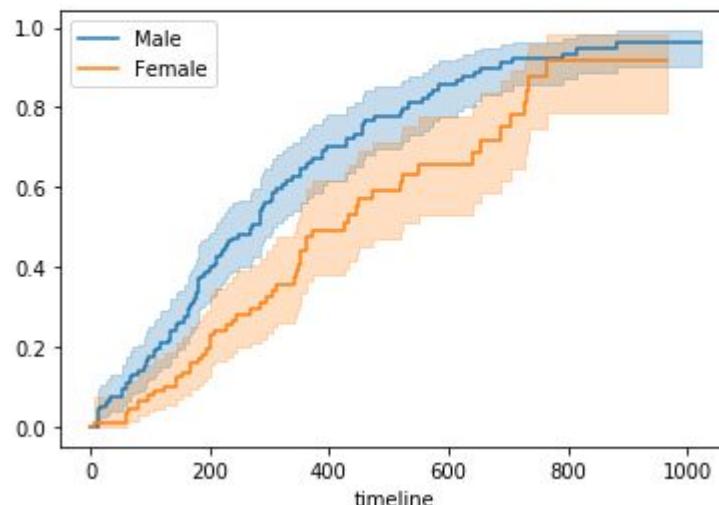


Upside-Down plots!

<https://www.kdnuggets.com/2020/07/guide-survival-analysis-python-part-3.html>

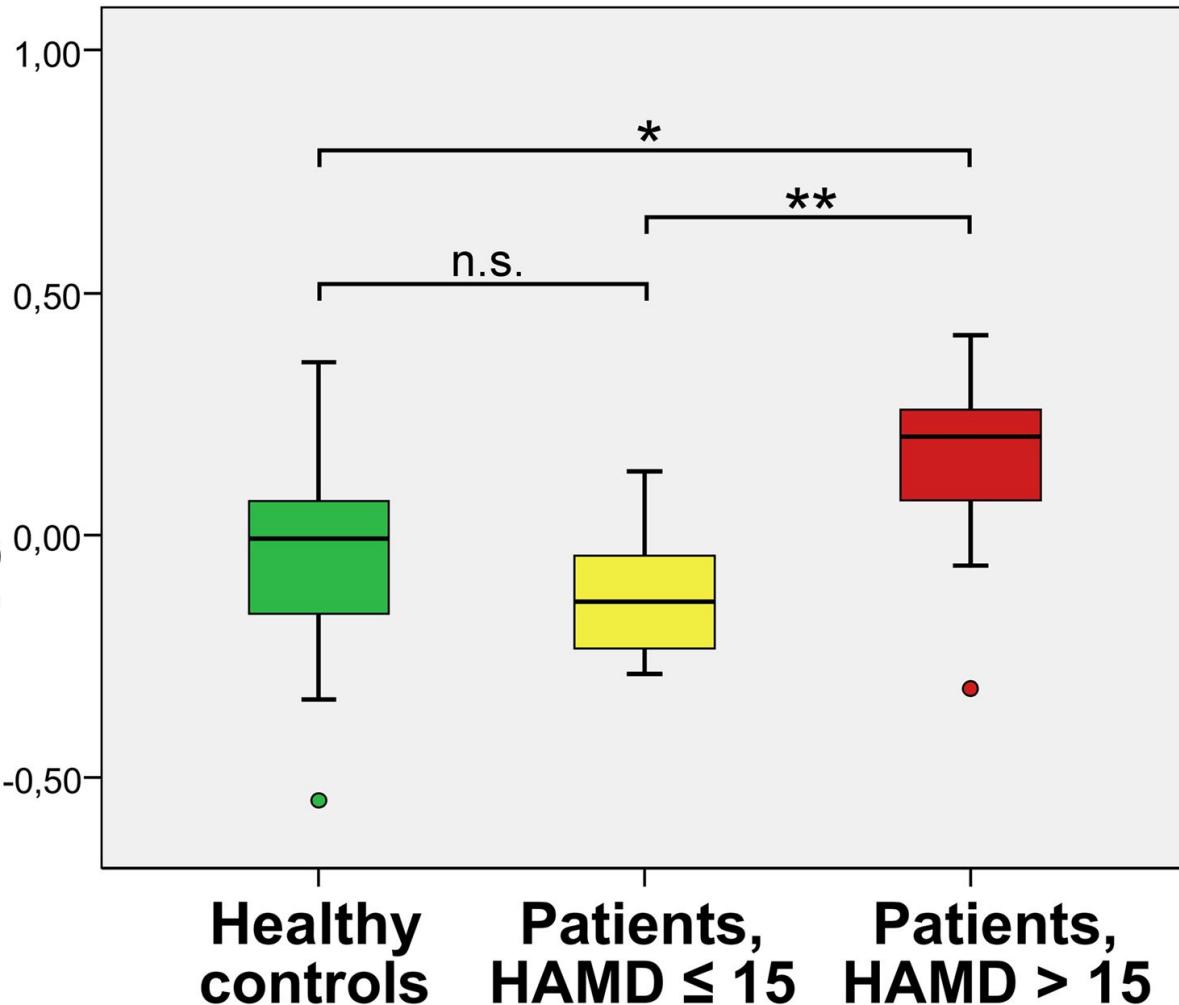
```
kmf_m.plot_cumulative_density()  
kmf_f.plot_cumulative_density()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x61cee55e08>
```



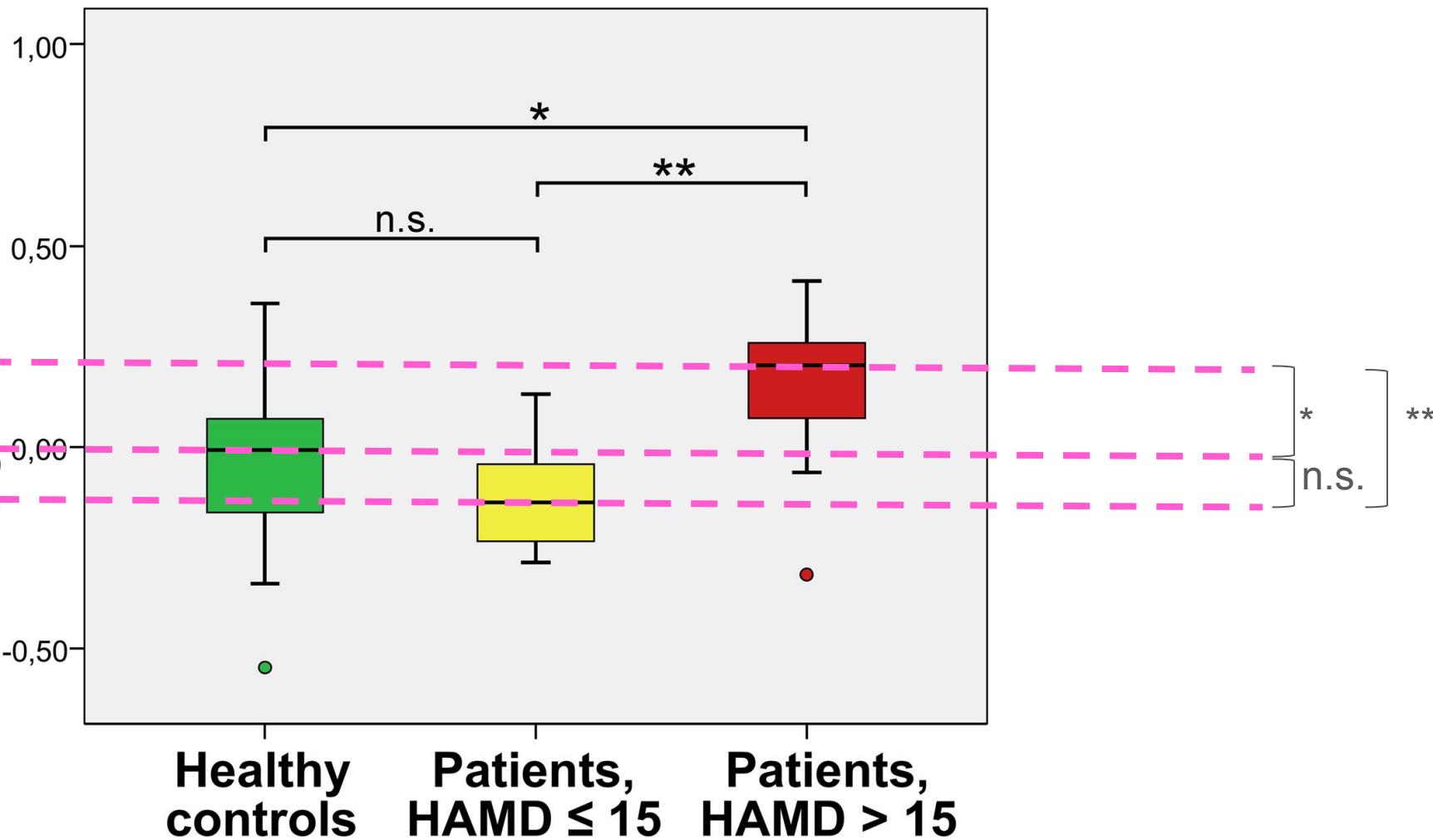
4 - COMPARING CURVES

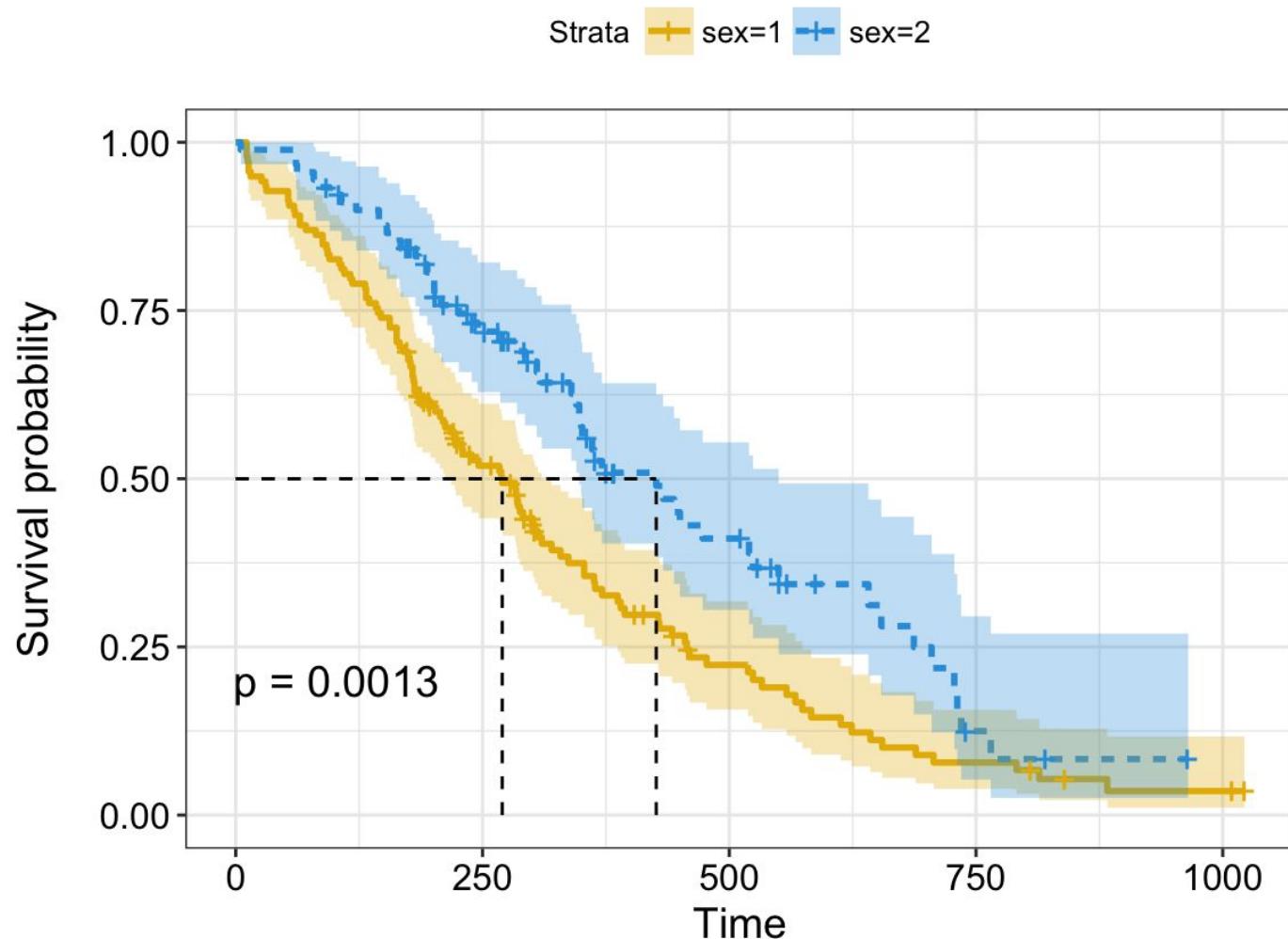
Functional connectivity of pgACC and AI



[https://ww2.mathworks.cn/
matlabcentral/answers/177
861-how-can-i-indicate-sig
nificance-in-boxplots?s_tid
=srchtile](https://ww2.mathworks.cn/matlabcentral/answers/177861-how-can-i-indicate-significance-in-boxplots?s_tid=srchtile)

**Functional connectivity
of pgACC and AI**





Source: NCCTG
lung cancer data set

Kaplan-Meier



Compare if **events' frequency** in both group
is similar to random chance

Exponential



Weibull

Compare if the **distribution's parameters** in groups
are similar to each others

Log-Normal



Others

Compare if the **hazard ratios** between groups
is different

Kaplan-Meier



Log-Rank

Exponential



Weibull

Parametric

Log-Normal

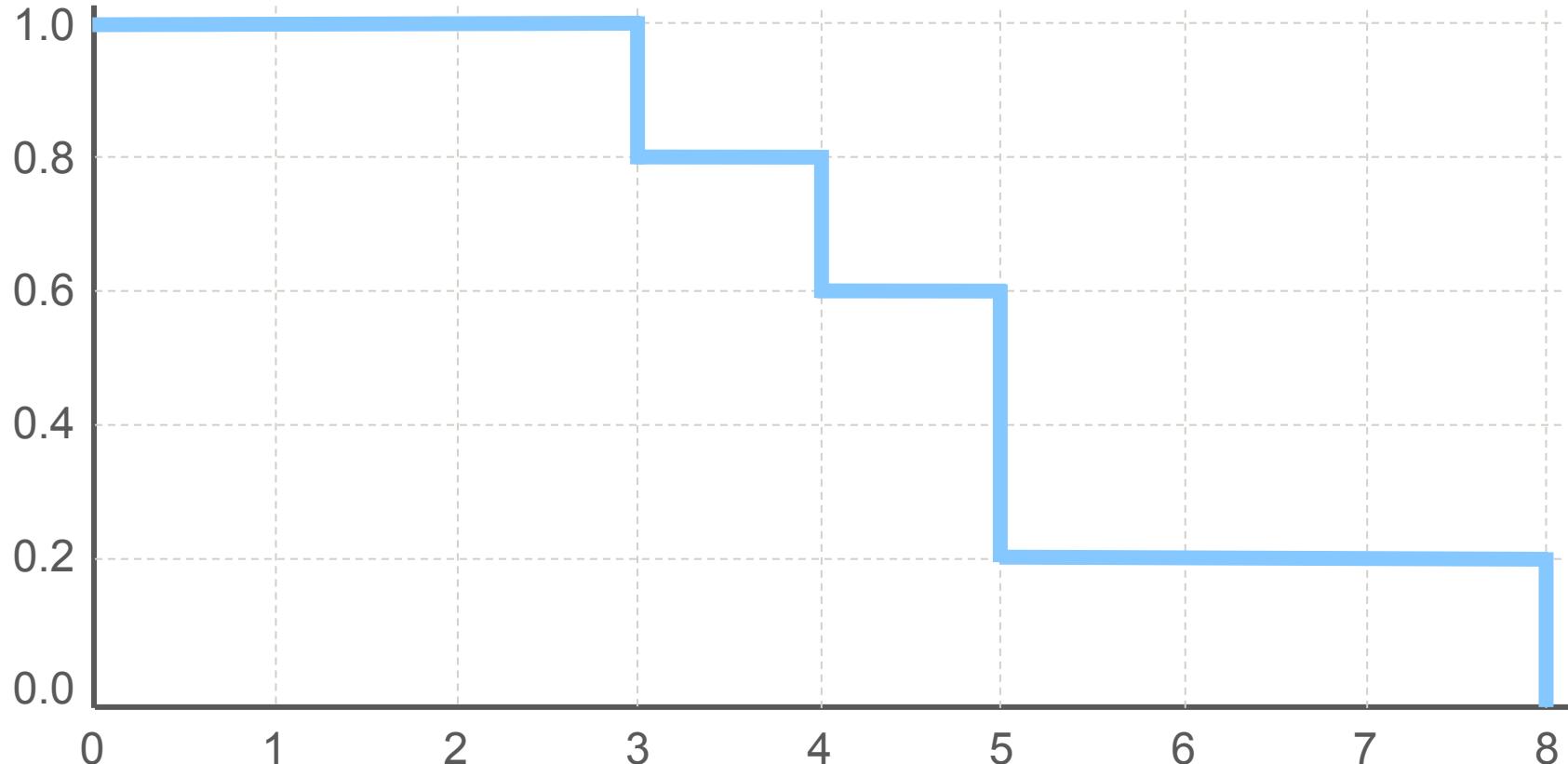


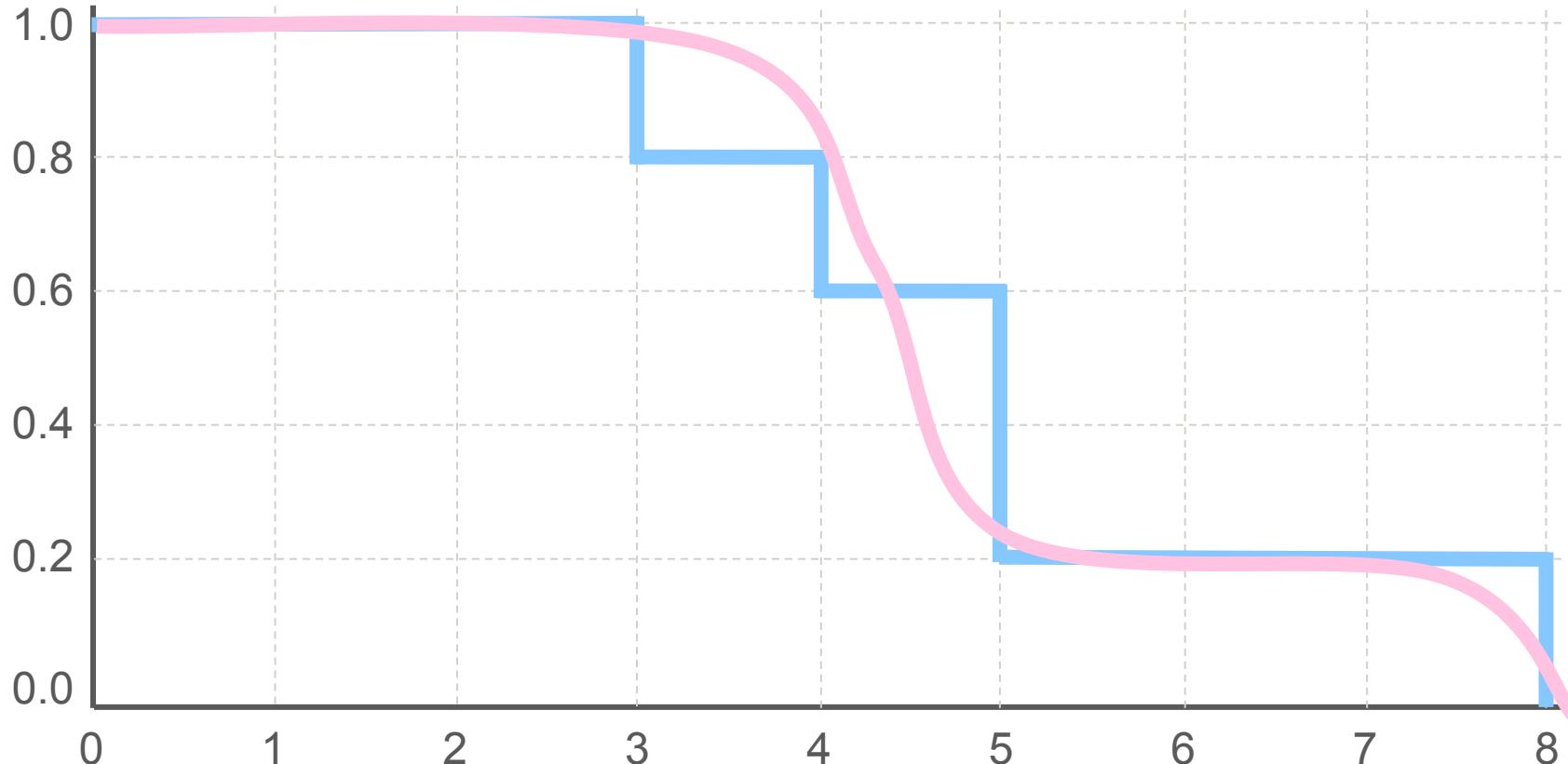
Others

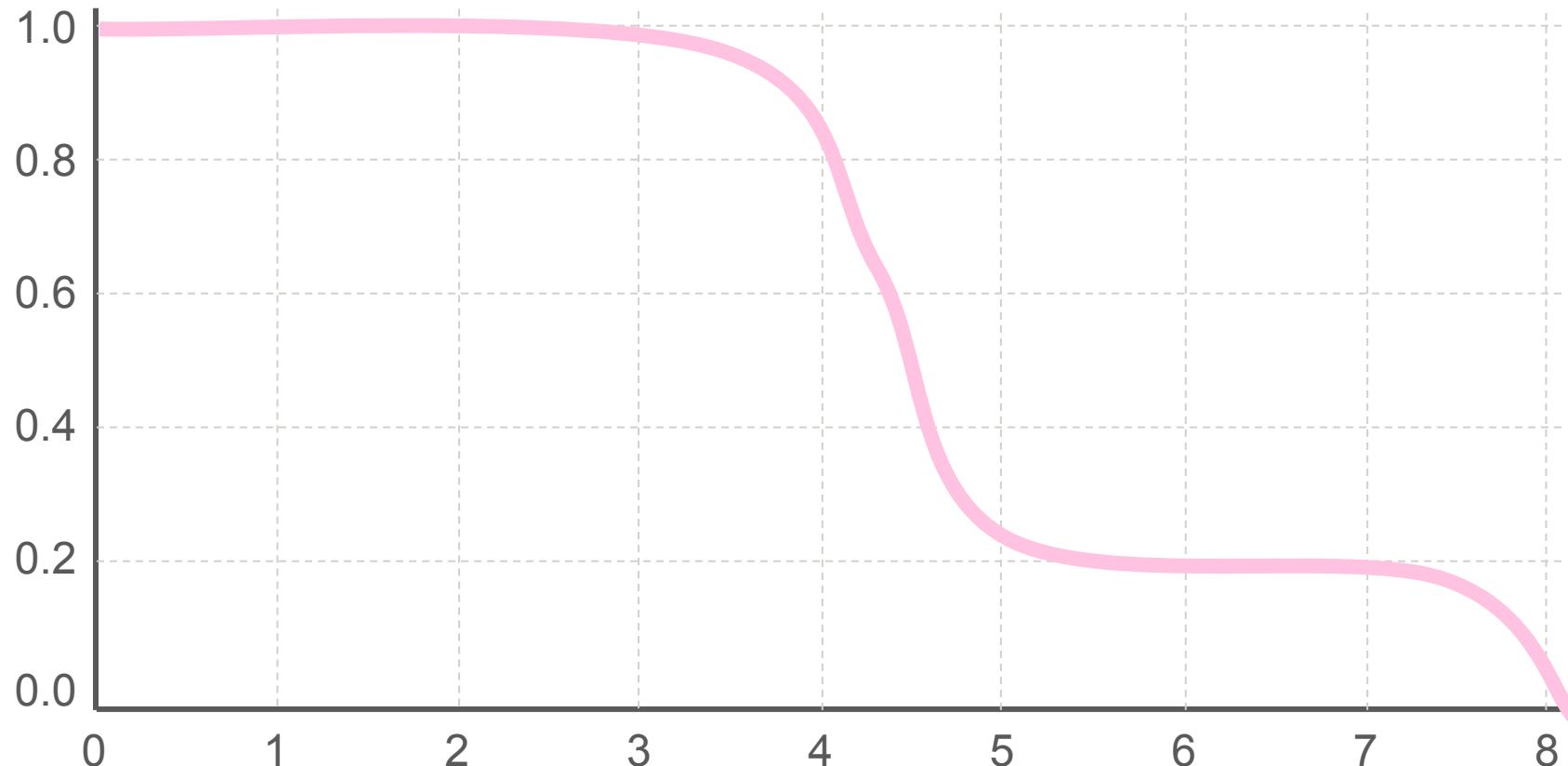
Cox's Proportional Hazard

f()

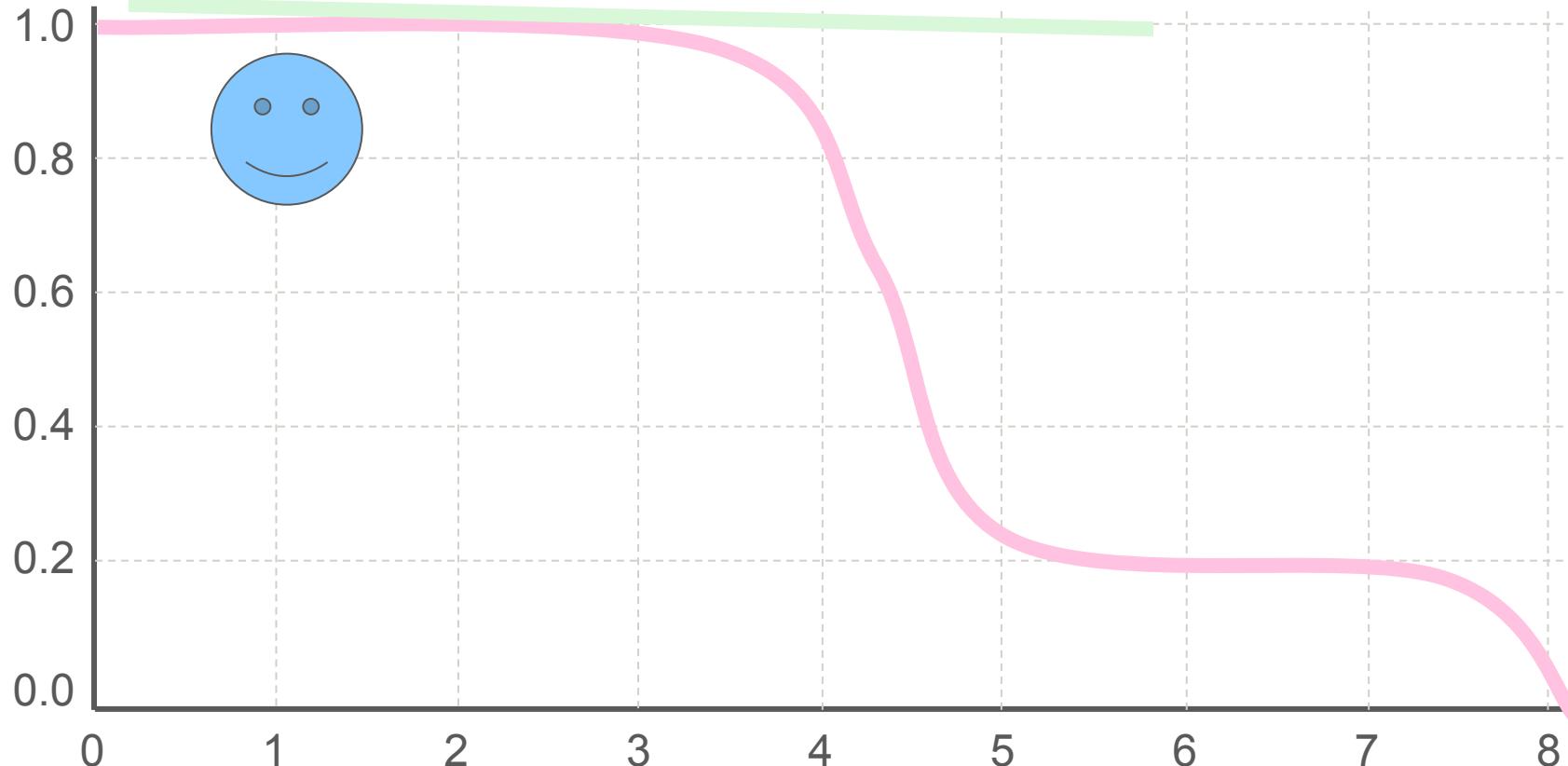


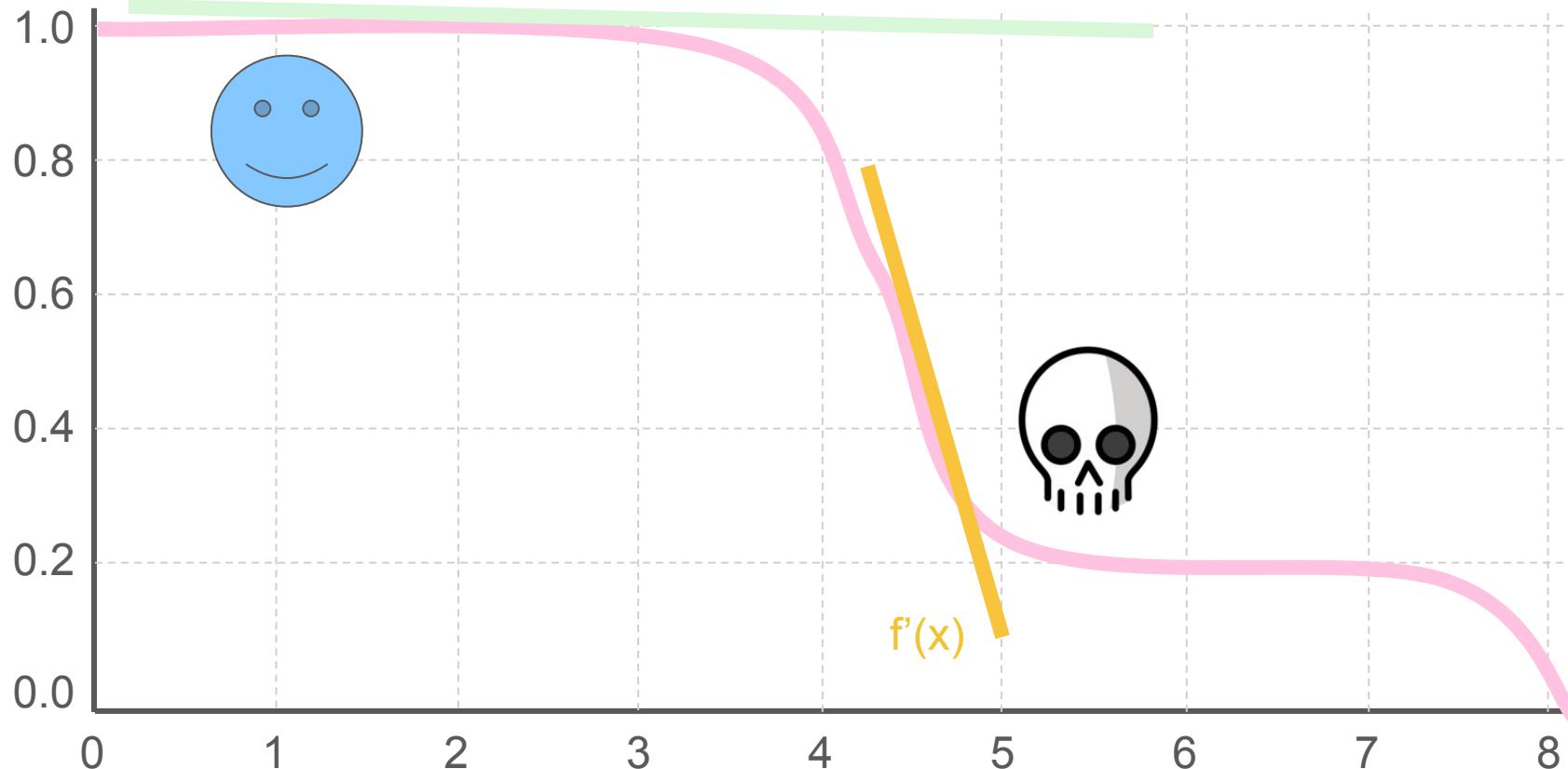


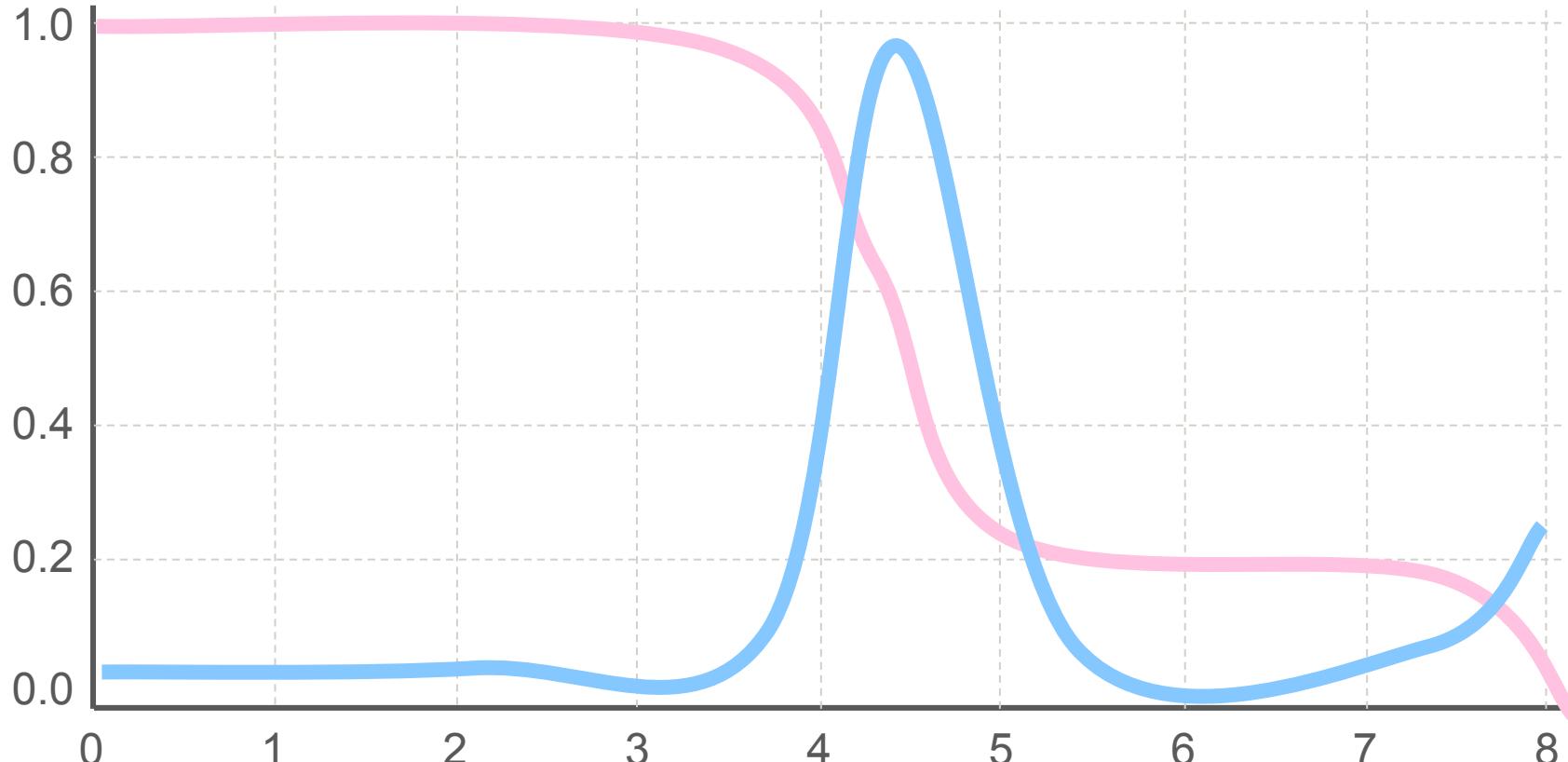




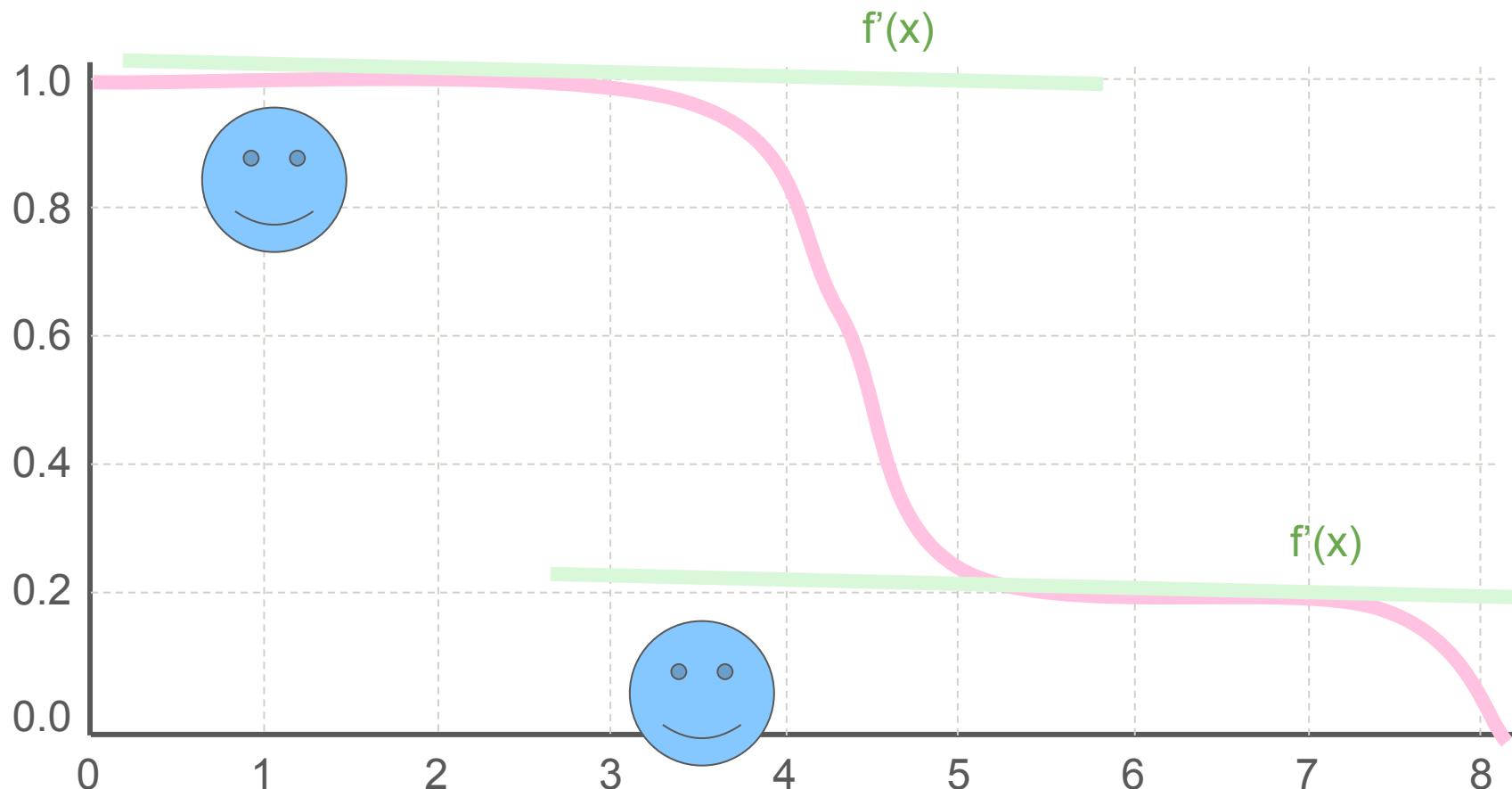
$f'(x)$

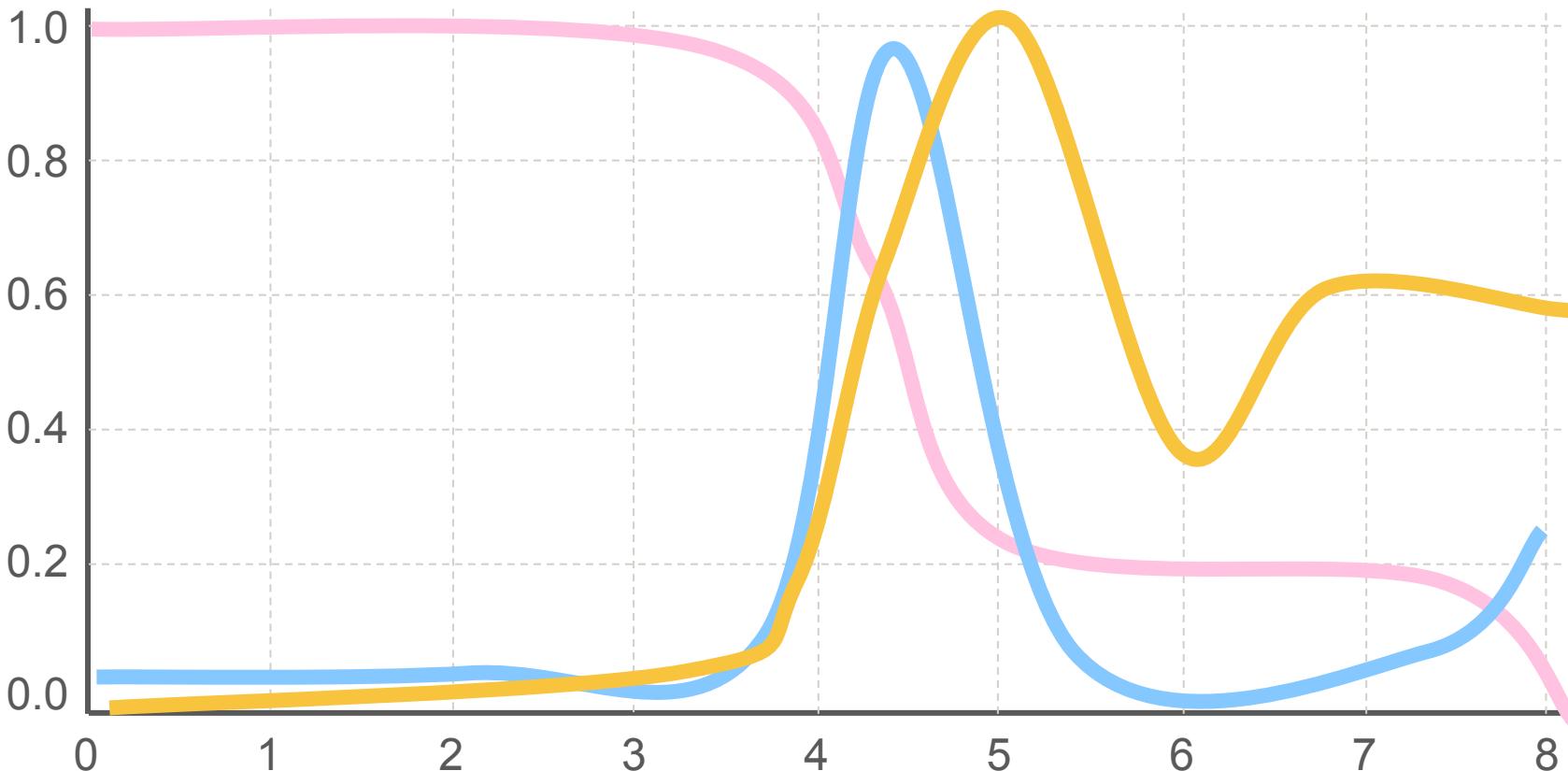


$f'(x)$ 



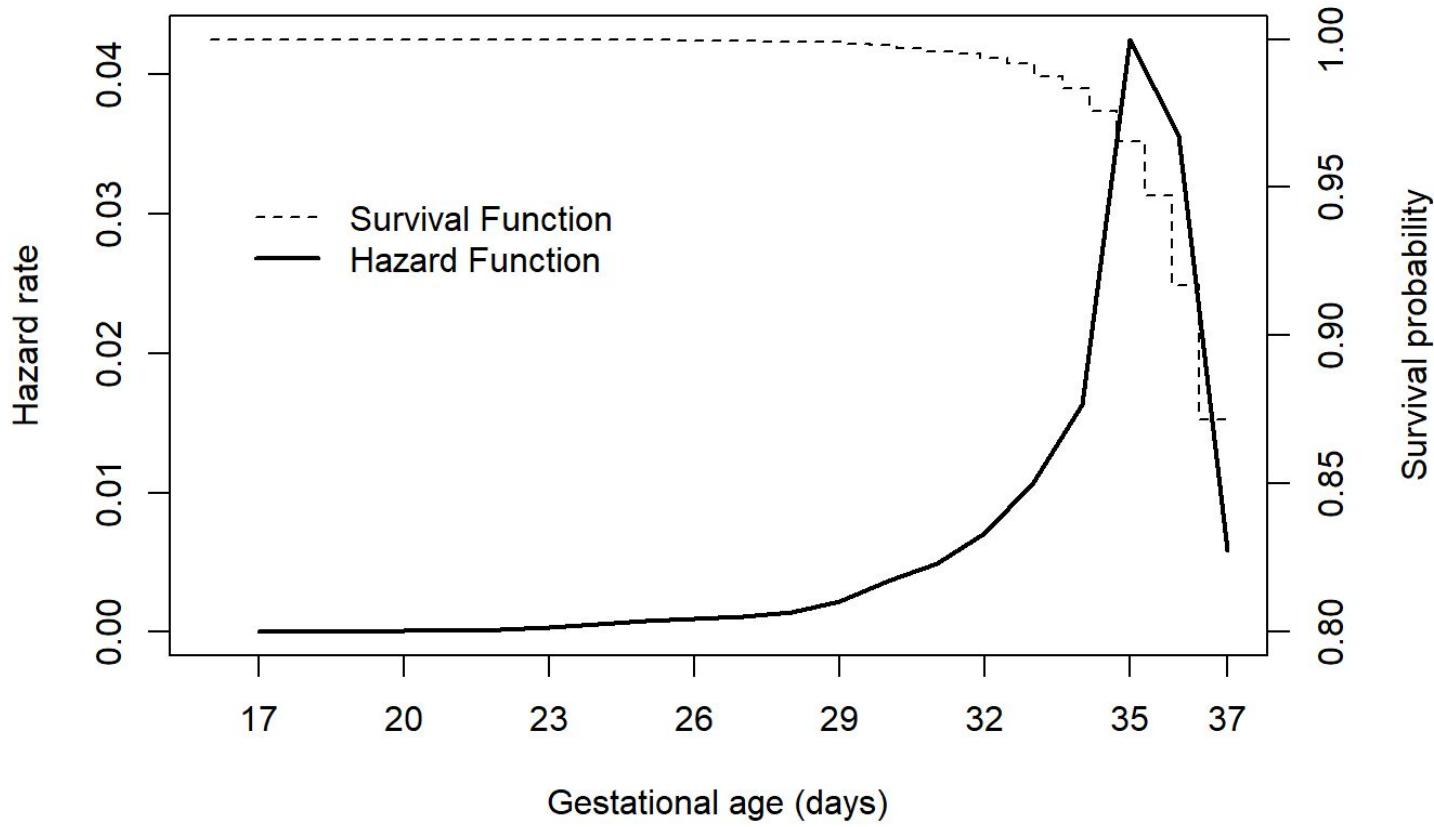
Hazard function?



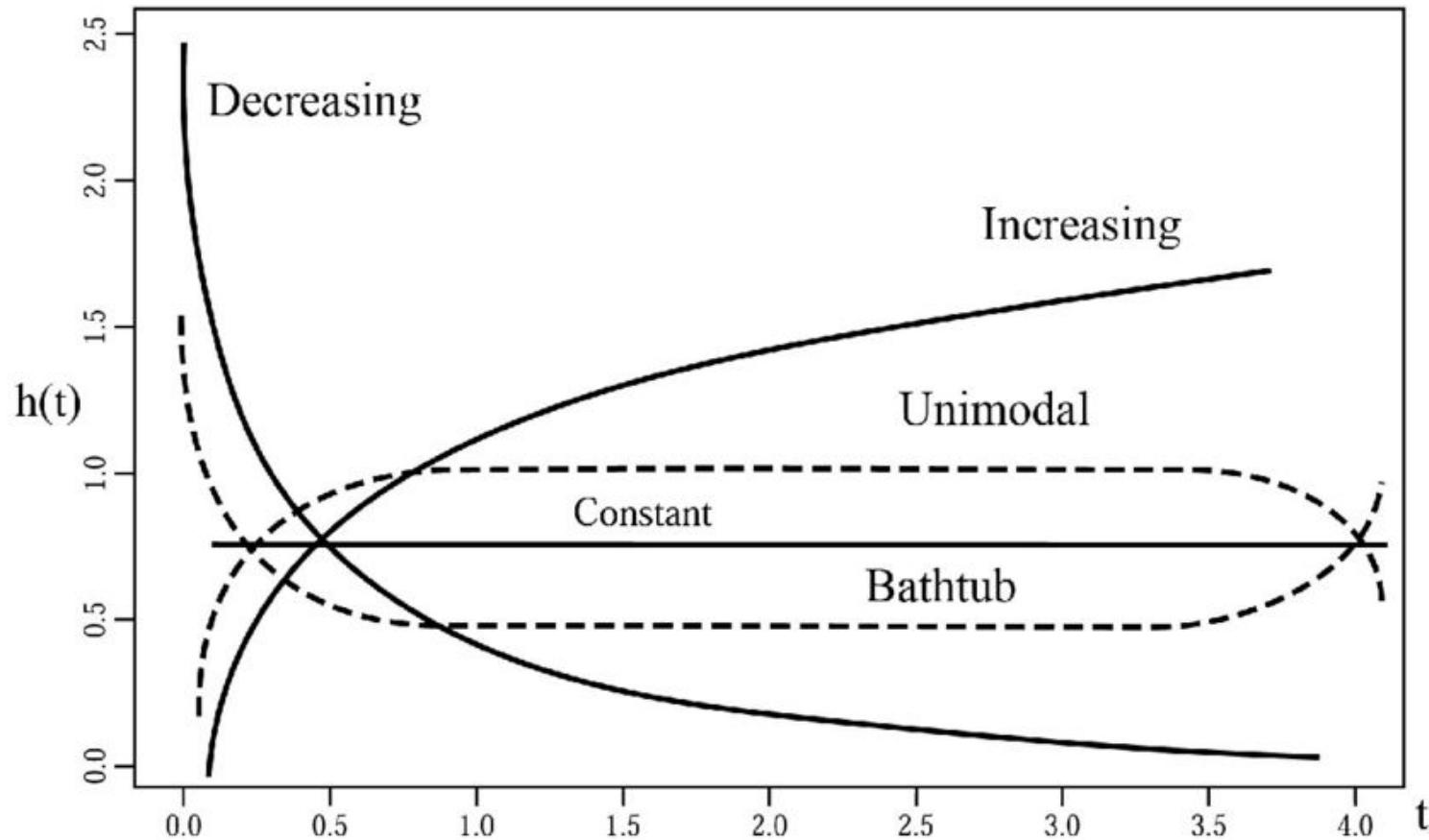


$$h(t) = \lim_{dt \rightarrow 0} \frac{\Pr(t \leq T < t + dt)}{dt \cdot S(t)} = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)}.$$

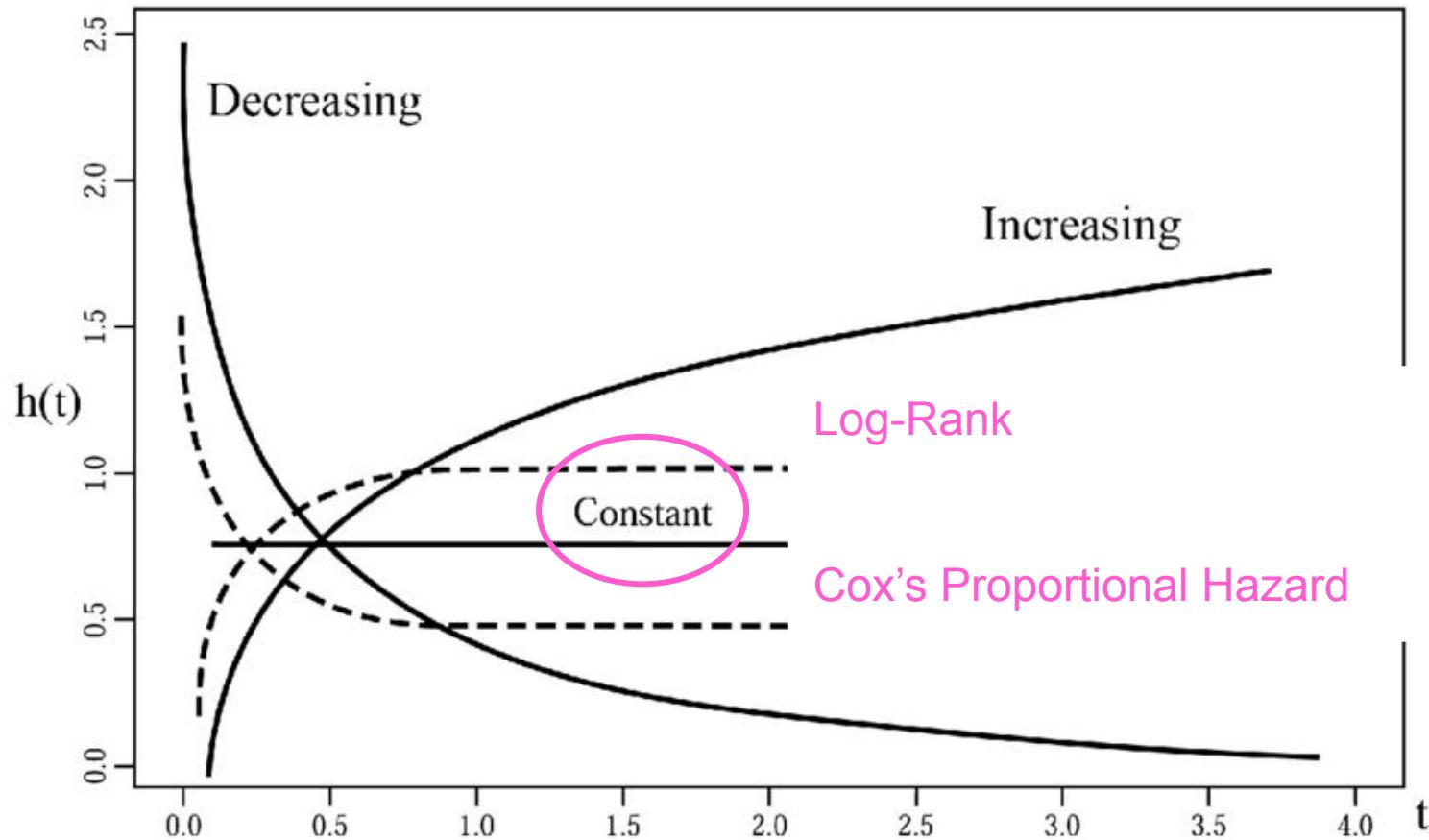
Hazard function



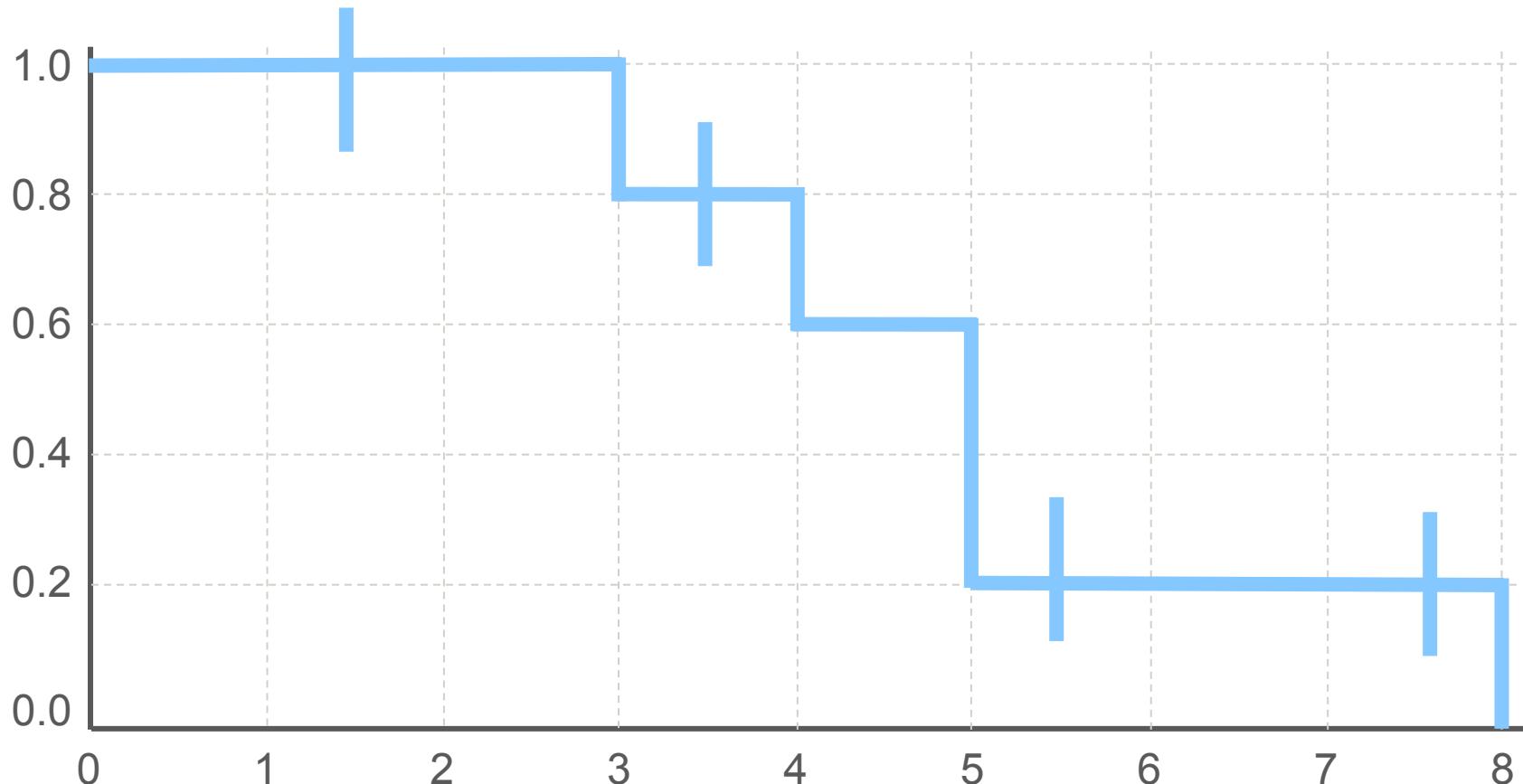
National Health and
Nutrition Examination
Surveys (NHANES)

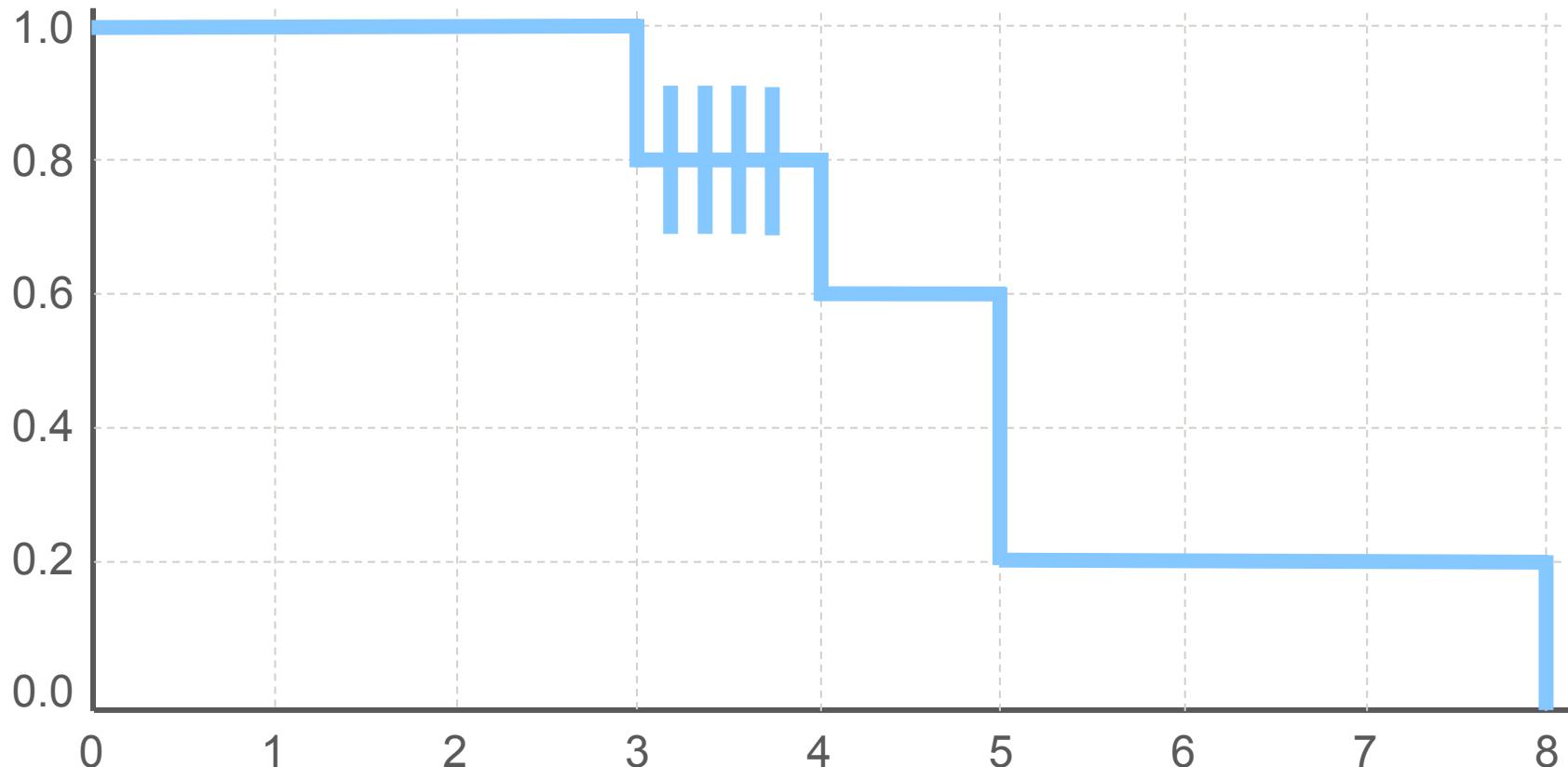


Bayesian and classical inference for the
generalized gamma distribution and
related models. 2018
Thesis for: p.h.D



Bayesian and classical inference for the
generalized gamma distribution and
related models. 2018
Thesis for: p.h.D









Log Rank Test

χ^2 - test

	Dead	Alive
Man		
Woman		

100

100

100

100

χ^2 - test

	Dead	Alive
Man	50	50
Woman	50	50

χ^2 - test

	Dead	Alive
Man	50	50
Woman	50	50

	Dead	Alive
Man	90	10
Woman	10	90

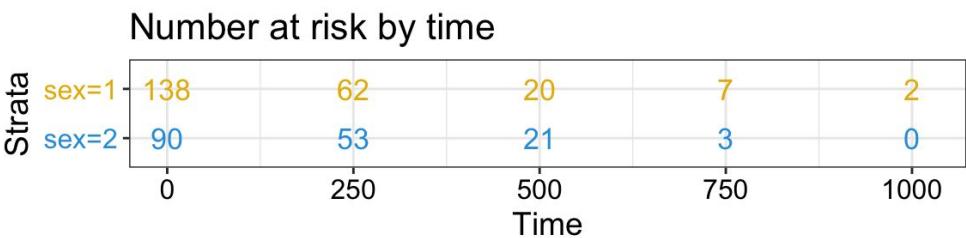
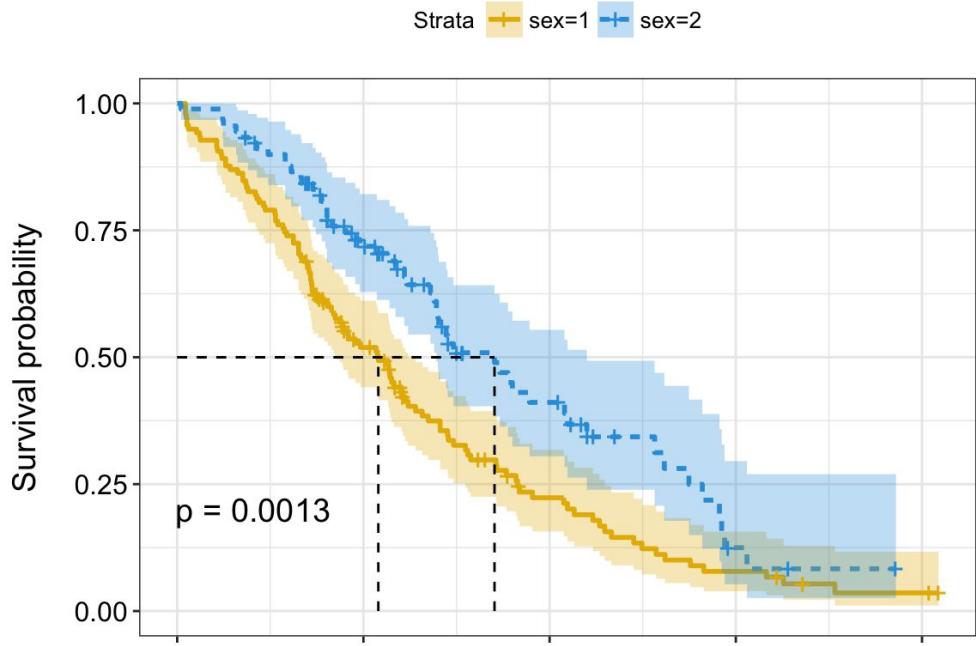
χ^2 - test

	Dead	Alive
Man	50	50
Woman	50	50

$$p = 1$$

	Dead	Alive
Man	90	10
Woman	10	90

$$p < 0.0001$$

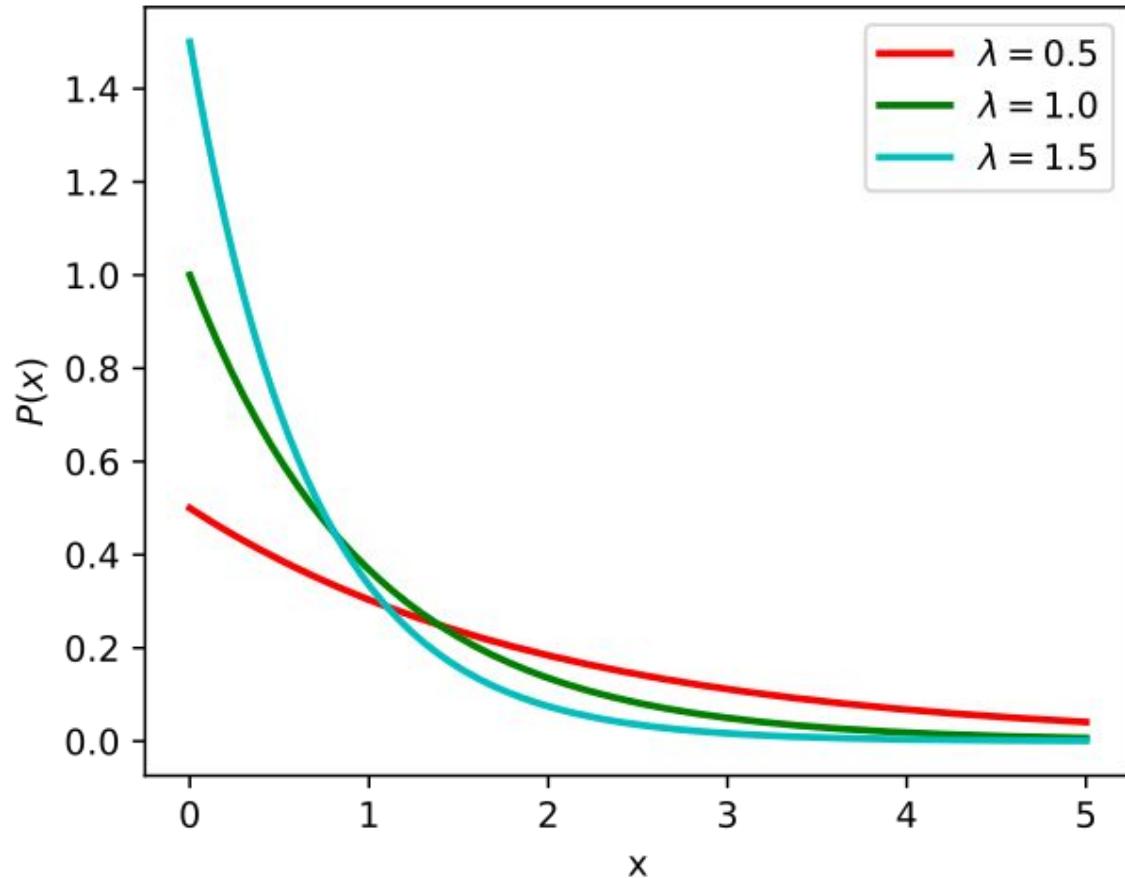


```

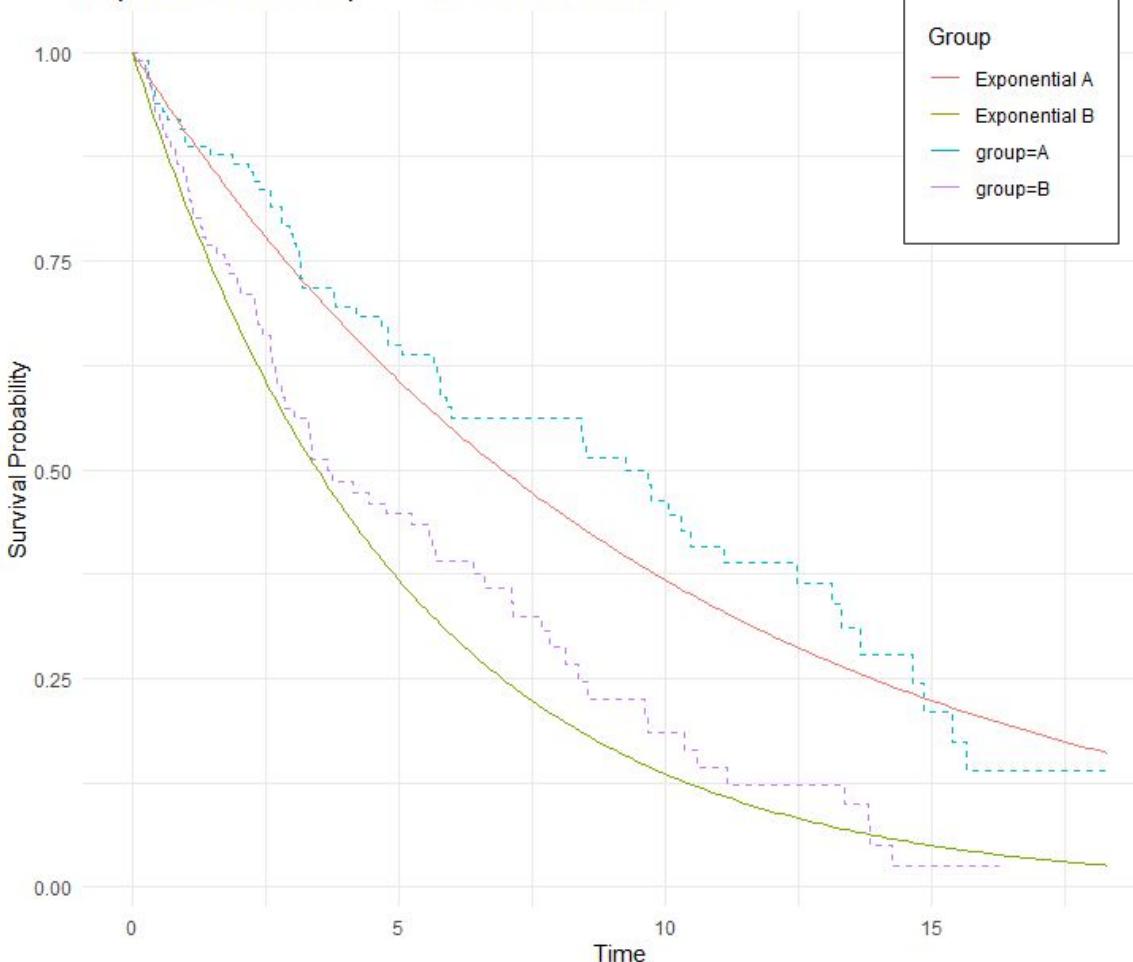
surv_diff <- survdiff(Surv(time, status) ~ sex, data = lung)
surv_diff
Call:
survdiff(formula = Surv(time, status) ~ sex, data = lung)
      N Observed Expected (0-E)^2/E (0-E)^2/V
sex=1 138       112     91.6    4.55    10.3
sex=2  90        53     73.4    5.68    10.3
Chisq= 10.3 on 1 degrees of freedom, p= 0.001

```

Parametric Test



Kaplan-Meier and Exponential Survival Curves



```
survreg(formula = Surv(time, status) ~ group, data = data, dist = "exponential")
```

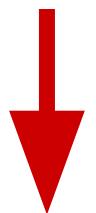
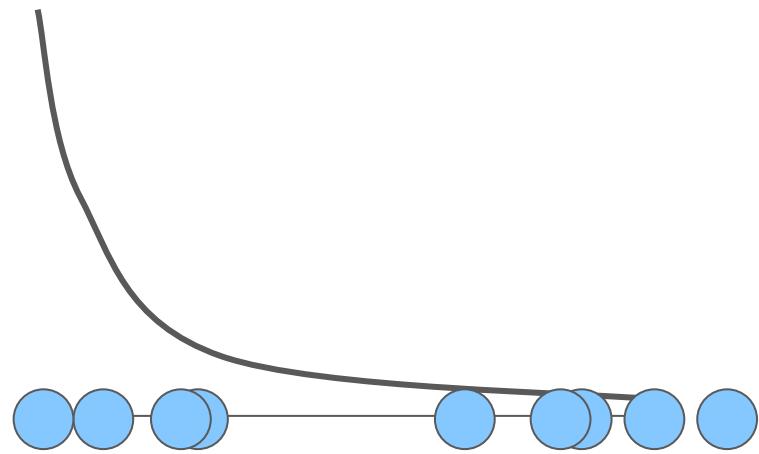
Coefficients:
(Intercept) groupB
2.4309254 -0.6648668

Scale fixed at 1

Loglik(model)=-392.6 Loglik(intercept only)=-399.6
Chisq= 14.06 on 1 degrees of freedom, **p= 0.000177**
n= 200

```
> cat("Likelihood Ratio Test Statistic:", lrt_stat, "\n")  
Likelihood Ratio Test Statistic: 2.273737e-13
```

```
> cat("P-value:", lrt_pvalue, "\n")  
P-value: 0
```



Π Probability

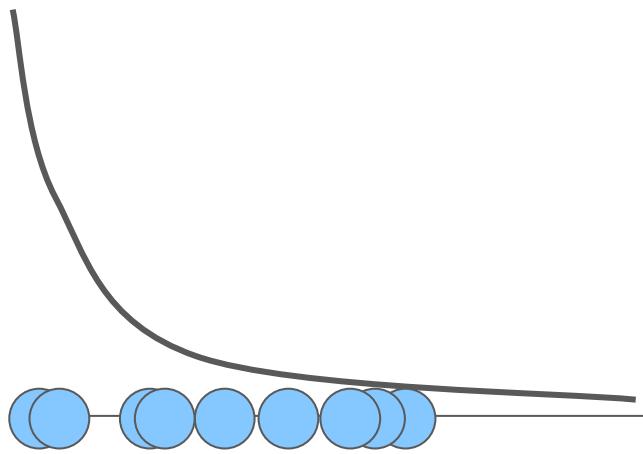


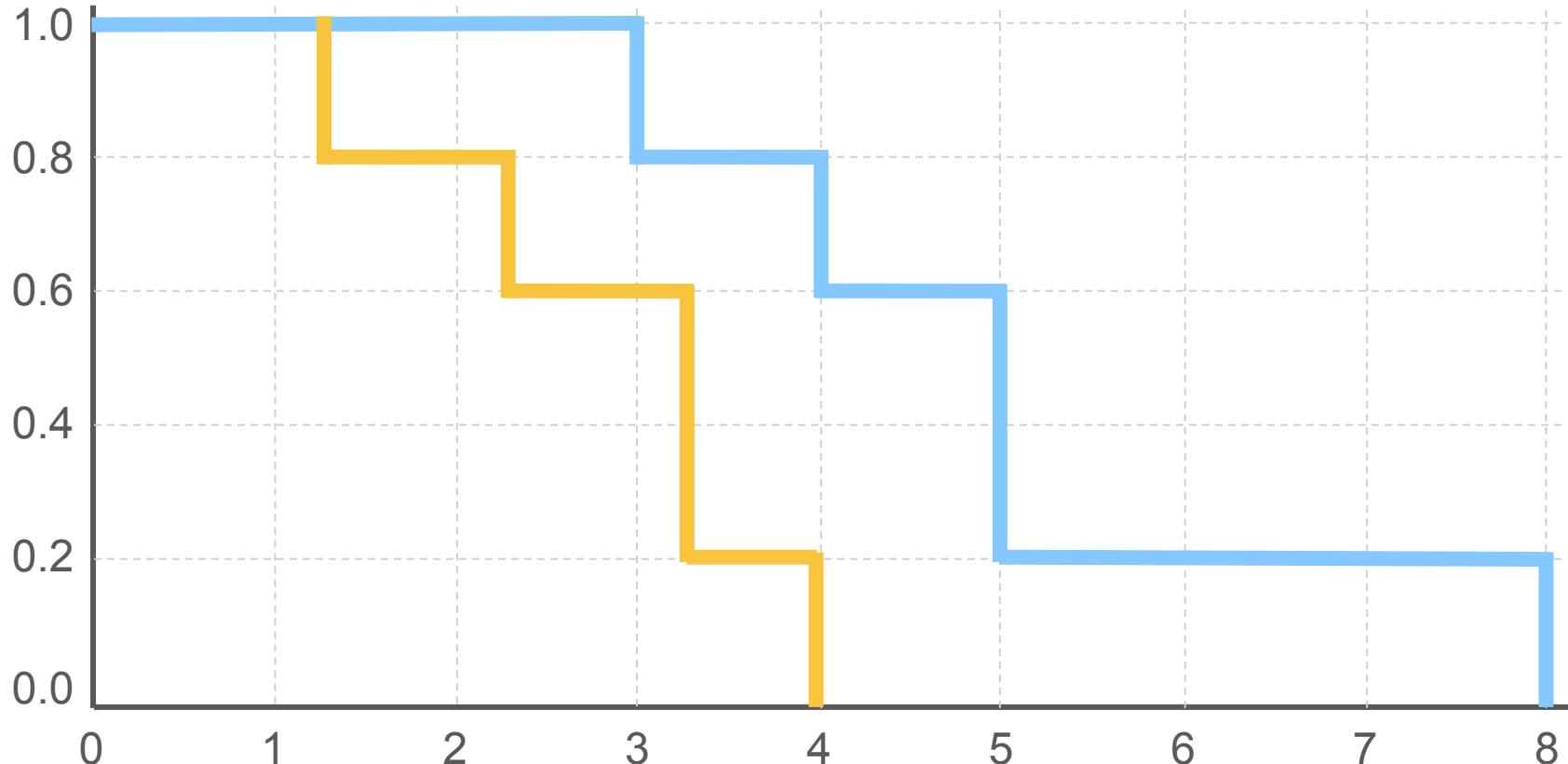
Table 1. Likelihood ratio (LR) test and standardised estimates of survival on young and old-growth juniper trees.

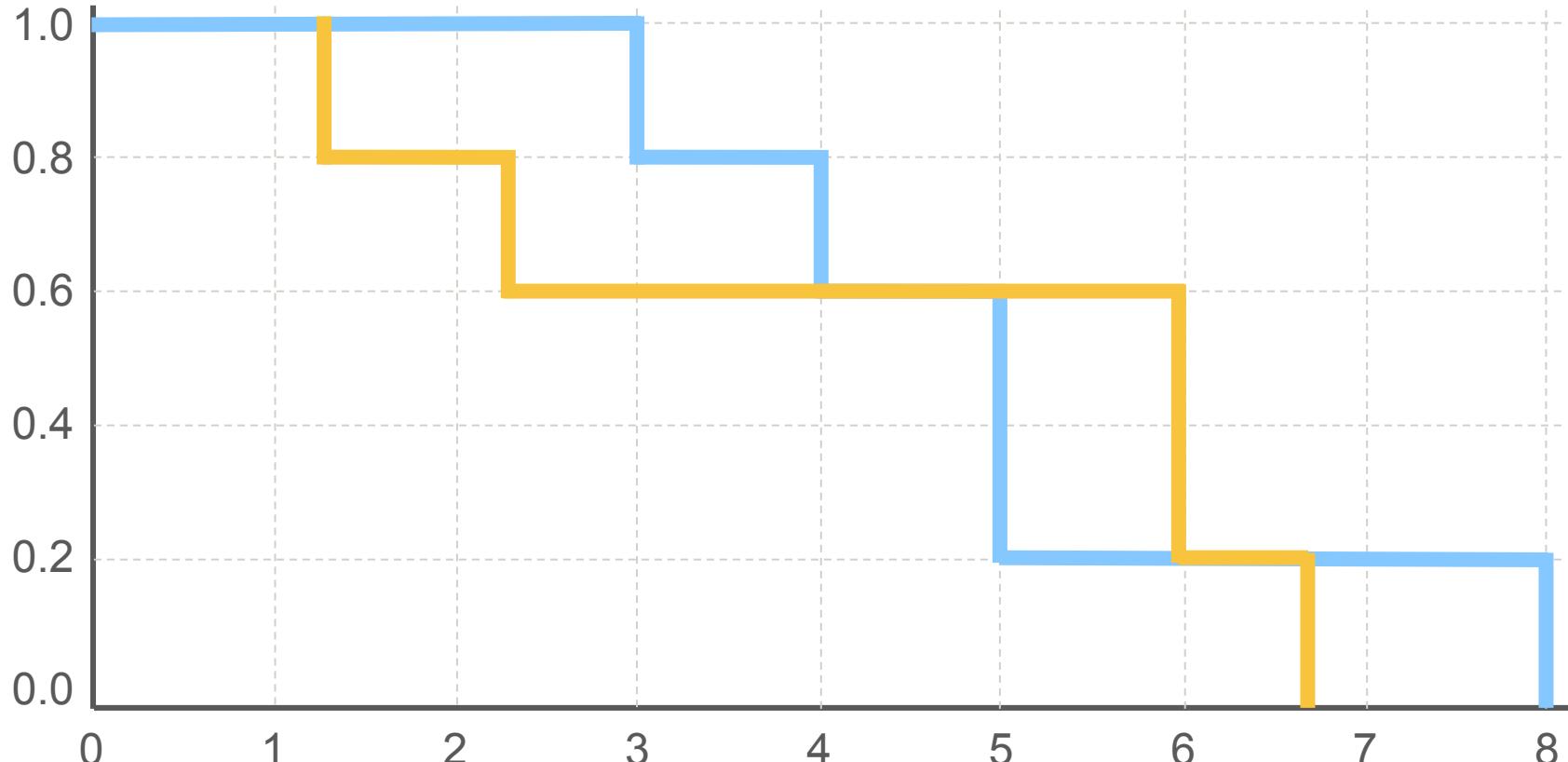
Variables	β	LR- χ^2	d.f.	P
Tree age	0.40	1.566	1	0.211
Lepidoptera species		10.069	1	0.002
Tree species		4.918	1	0.027
Tree age \times Lepidoptera species		1.997	1	0.158
Tree age \times tree species		1.566	1	0.211
Lepidoptera species \times tree species		4.182	1	0.041
Tree age \times Lepidoptera species \times tree species		11.611	1	0.001

Results from a type III analysis of deviance from a logistic generalised linear model with a binomial error distribution to model frequency of survival for *Callophrys gryneus* and *Glena quinquellearia* on young and old *Juniperus osteosperma* and *Juniperus occidentalis* trees. β represents the standardised beta coefficient. A significant predictor of the frequency is indicated in bold ($P < 0.05$).

Preference and performance of Lepidoptera varies with tree age in juniper woodlands
2018 Ecological Entomology 44(1)
DOI:10.1111/een.12682

Cox's Proportional Hazard





$$\beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \dots$$

$$\lambda(t)$$

$\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots$

$\lambda(t)$

Coefficients

Covariates

$\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots$

$\lambda(t)$

$3\text{Smoking} + -7\text{Sex} + 0.1\text{BMI} + \dots$

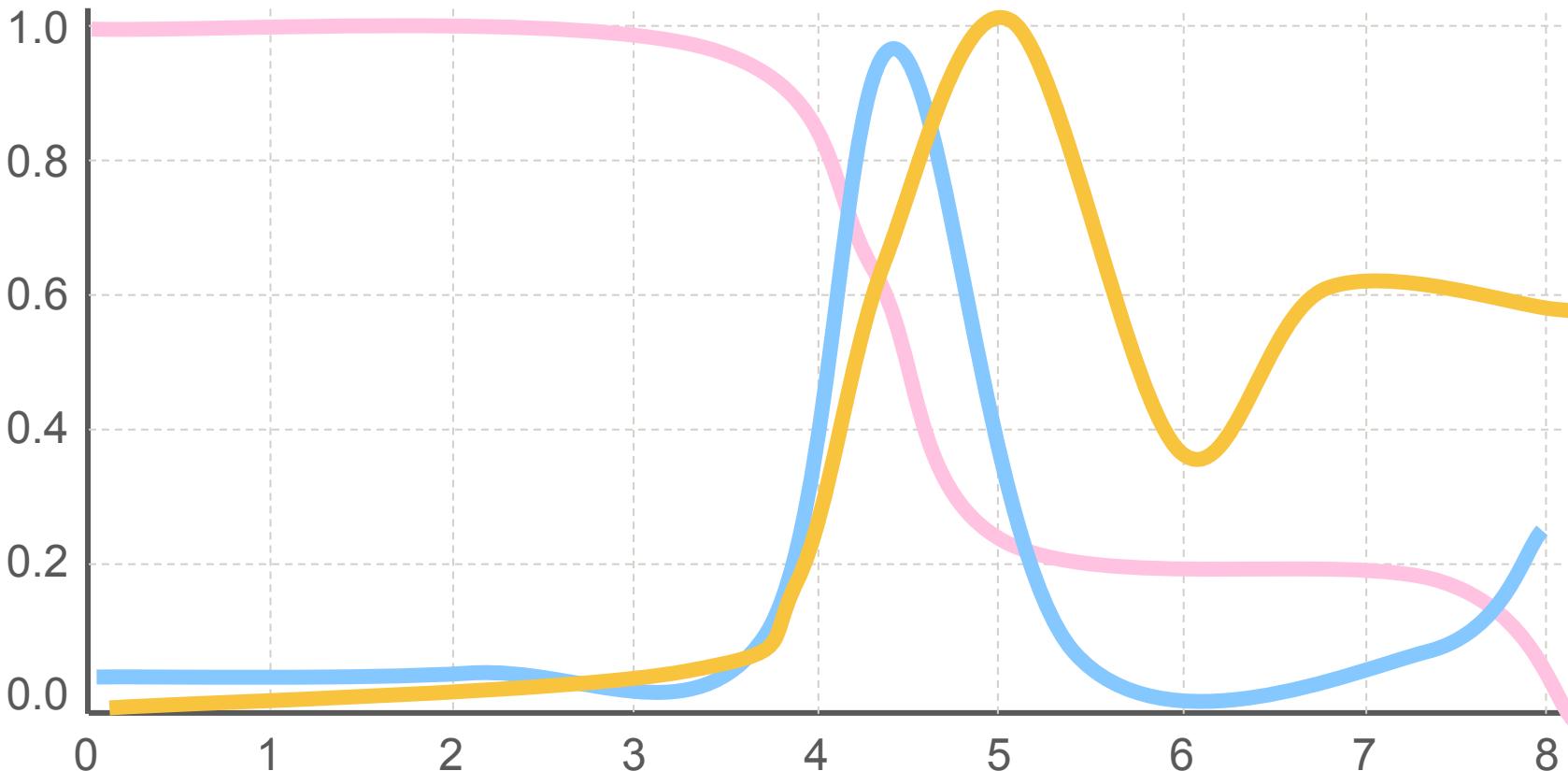
$\lambda(t)$

Covariant	Coefficient	Hazard	Effect
Smoking	3		Increase ↑
Sex	-7		Decrease ↓
BMI	0.1		Meh... ↗

0

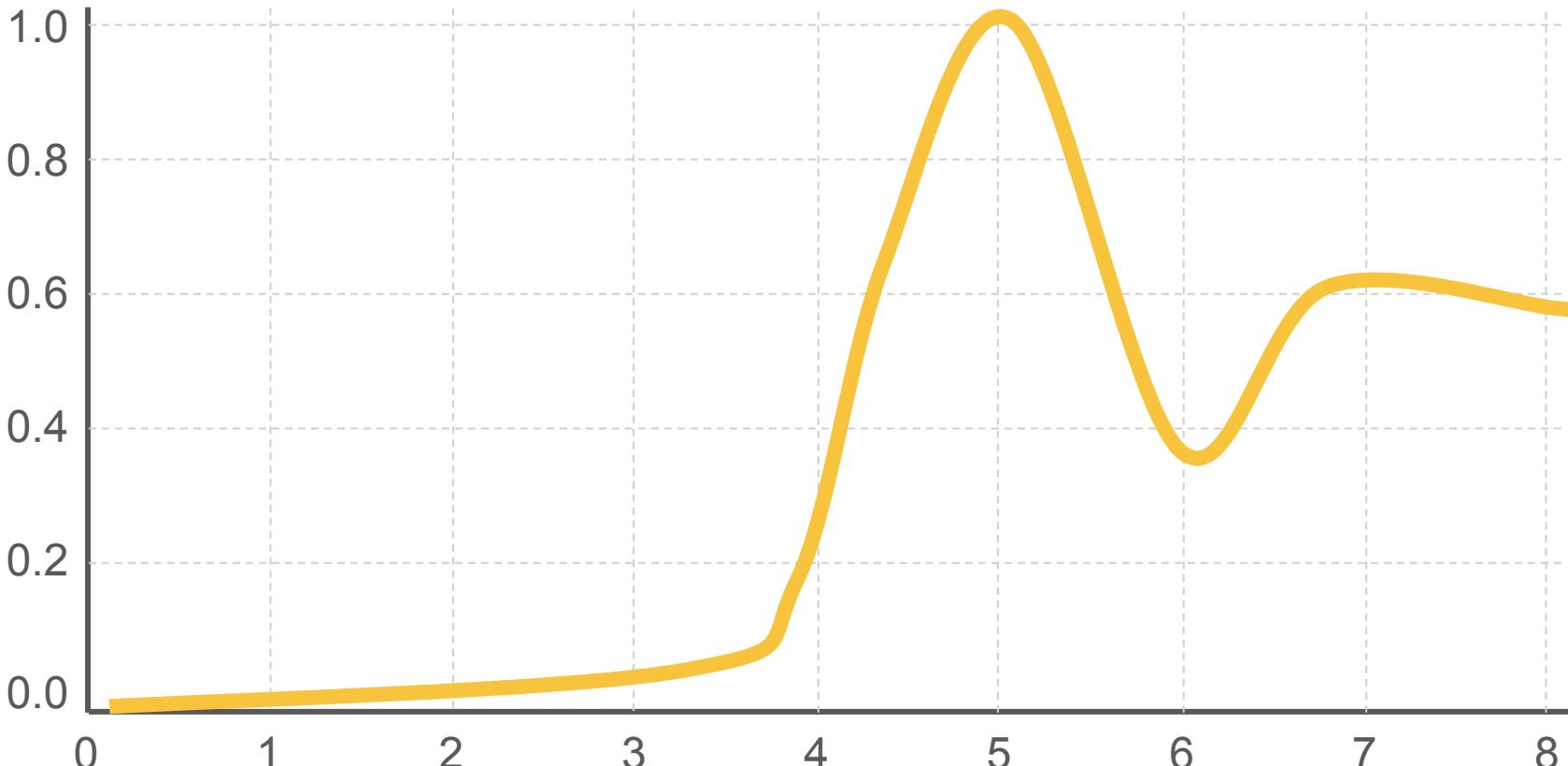
1

Covariant	Coefficient	Hazard	Effect
Smoking	3	1.7	Increase ↑
Sex	-7	0.2	Decrease ↓
BMI	0.1	1.05	Meh... ↗



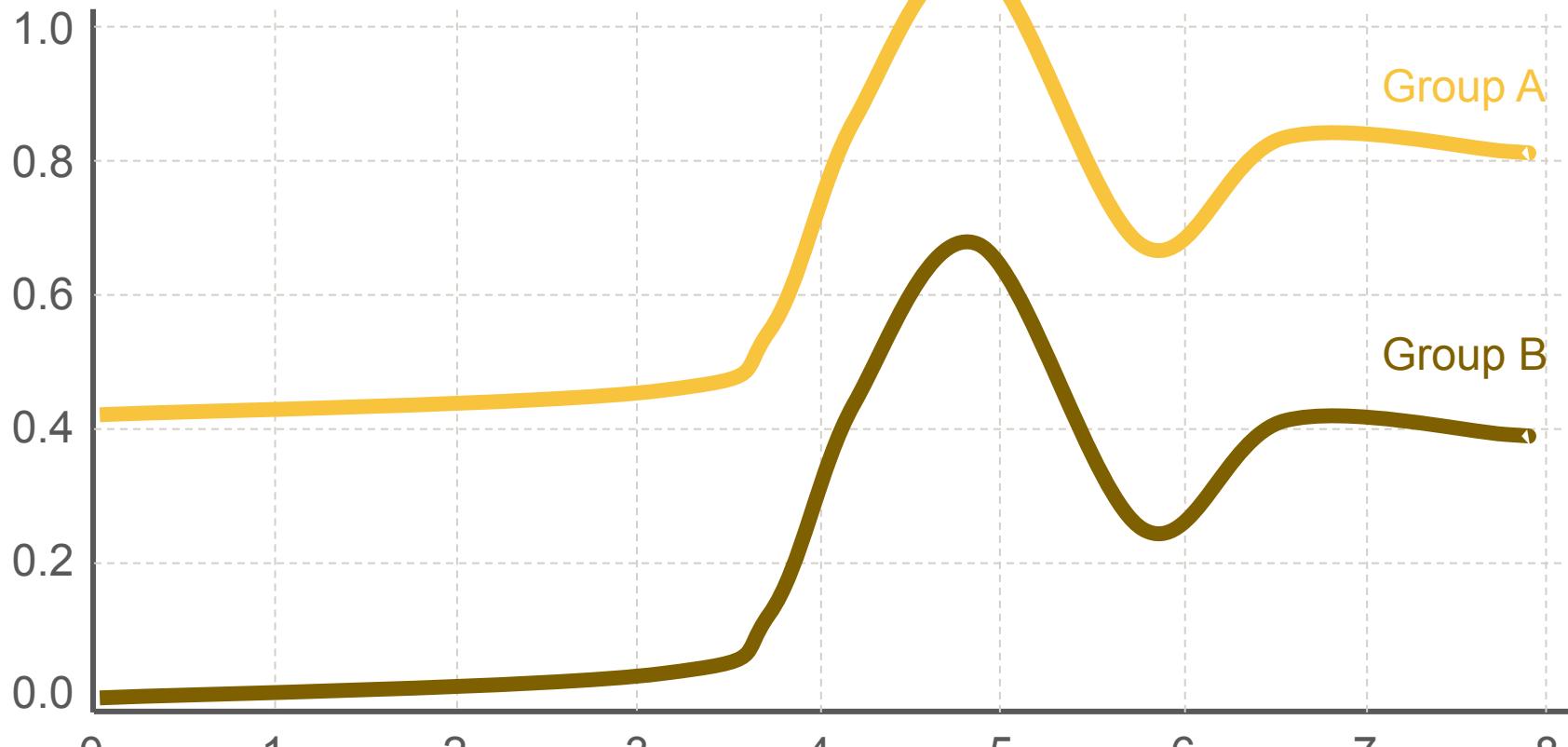
$$h(t) = \lim_{dt \rightarrow 0} \frac{\Pr(t \leq T < t + dt)}{dt \cdot S(t)} = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)}.$$

Hazard function

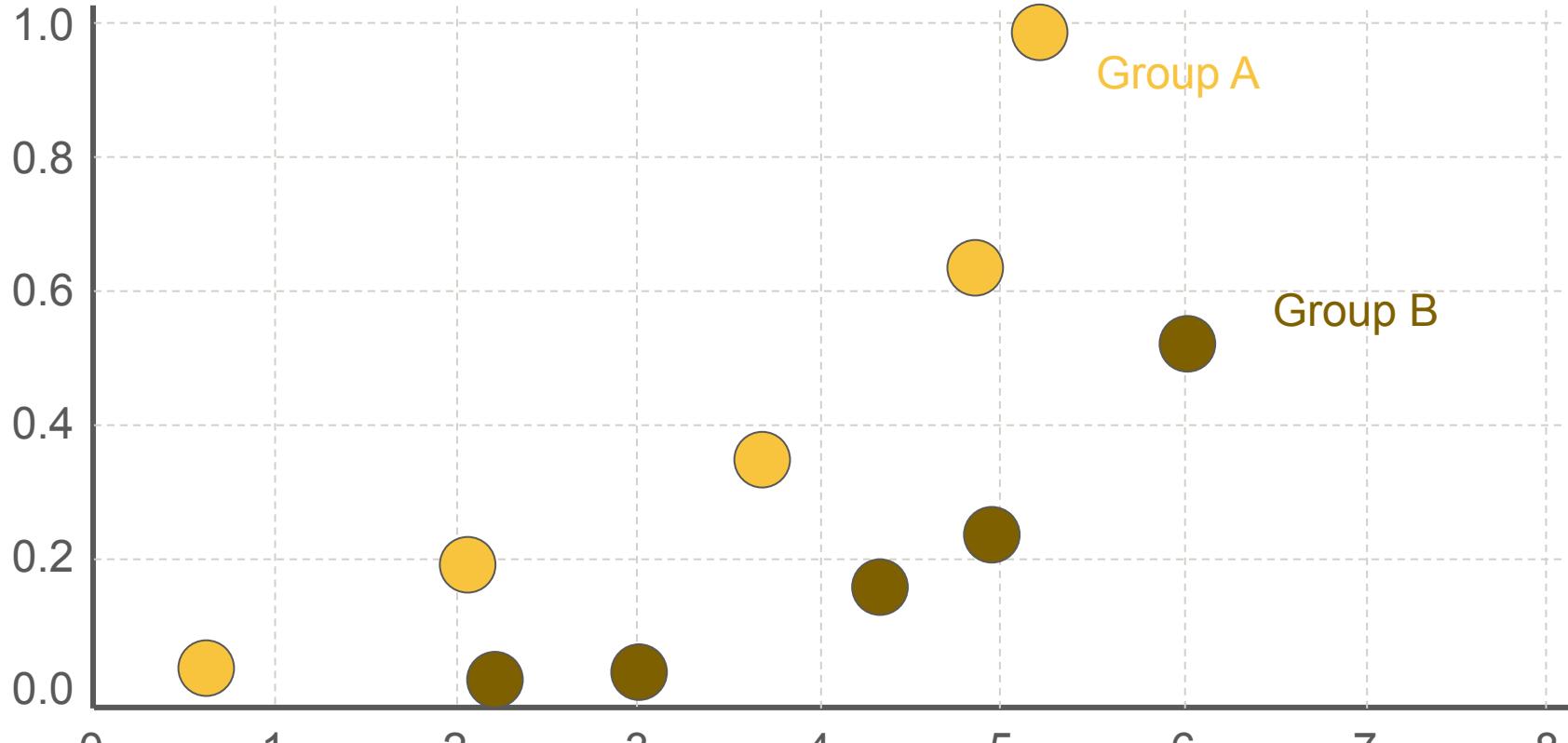


$$h(t) = \lim_{dt \rightarrow 0} \frac{\Pr(t \leq T < t + dt)}{dt \cdot S(t)} = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)}.$$

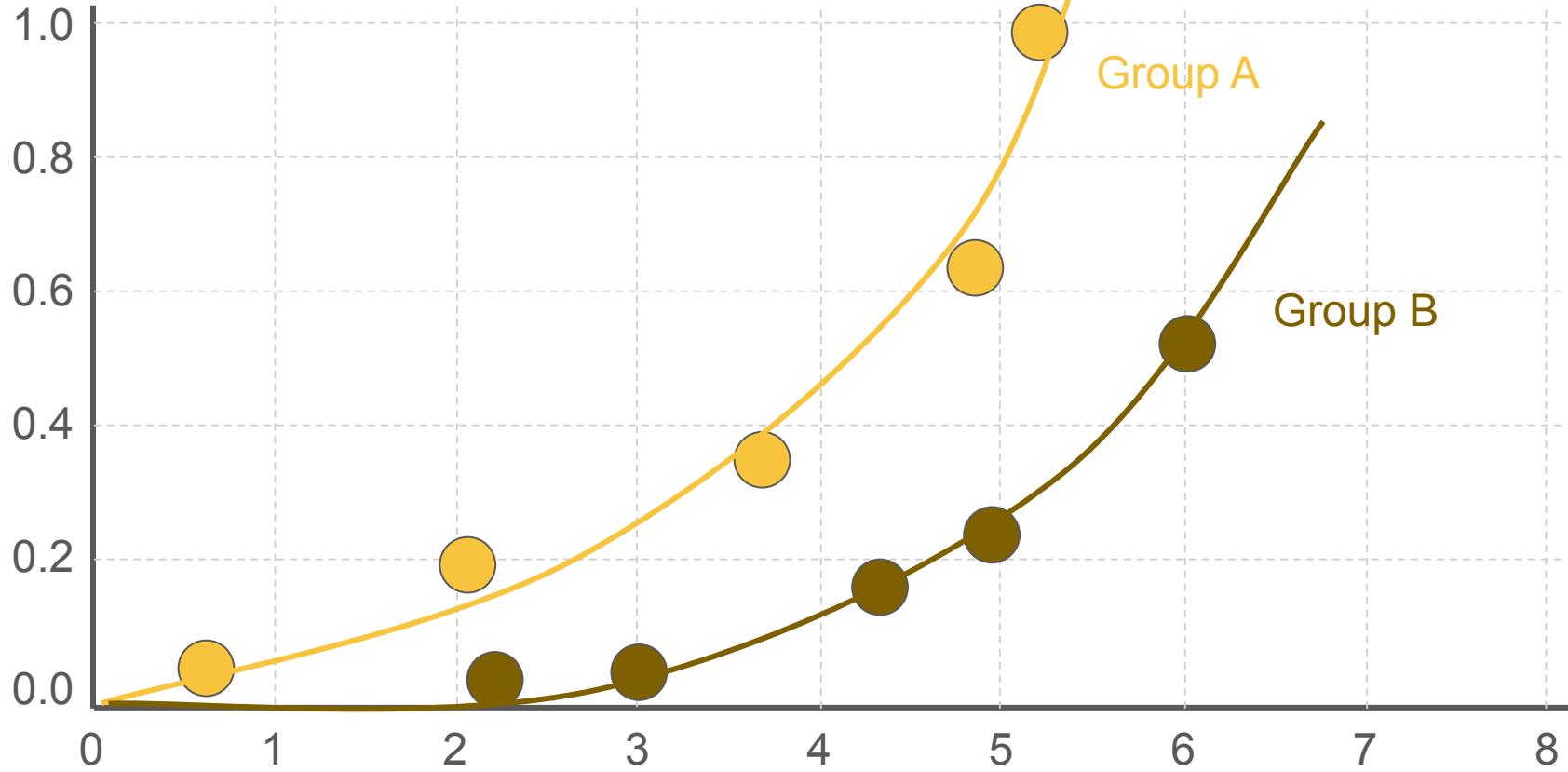
Hazard function



Hazard functions



Hazard functions



Hazard functions

```
## Call:  
## coxph(formula = Surv(futime, event) ~ gender + age_years + source +  
##         days_onset_hosp, data = linelist_surv)  
##  
##   n= 2772, number of events= 1180  
##   (1767 observations deleted due to missingness)  
##  
##             coef exp(coef)  se(coef)      z Pr(>|z|)  
## genderm      0.004710  1.004721  0.060827  0.077  0.9383  
## age_years     -0.002249  0.997753  0.002421 -0.929  0.3528  
## sourceother    0.178393  1.195295  0.084291  2.116  0.0343 *  
## days_onset_hosp -0.104063  0.901169  0.014245 -7.305 2.77e-13 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
##             exp(coef) exp(-coef) lower .95 upper .95  
## genderm       1.0047     0.9953    0.8918    1.1319  
## age_years      0.9978     1.0023    0.9930    1.0025  
## sourceother    1.1953     0.8366    1.0133    1.4100  
## days_onset_hosp 0.9012     1.1097    0.8764    0.9267  
##  
## Concordance= 0.566  (se = 0.009 )  
## Likelihood ratio test= 71.31  on 4 df,  p=1e-14  
## Wald test        = 59.22  on 4 df,  p=4e-12  
## Score (logrank) test = 59.54  on 4 df,  p=4e-12
```

```
## Call:  
## coxph(formula = Surv(futime, event) ~ gender + age_years + source +  
##         days_onset_hosp, data = linelist_surv)  
##  
##   n= 2772, number of events  
##   (1767 observations deleted)  
  
##  
##             Coefficient          Significant  
##  
##            coef exp(coef)    se(coef)      z Pr(>|z|)  
## genderm     0.004710  1.004721  0.060827  0.077  0.9383  
## age_years   0.002249  0.997753  0.002421 -0.929  0.3528  
## sourceother 0.178391  1.195295  0.084291  2.116  0.0343 *  
## days_onset_hosp -0.104063  0.901169  0.014245 -7.305 2.77e-13 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '  
##  
##             exp(coef) exp(-coef) lower .95 upper .95  
## genderm      1.0047     0.9953    0.8918    1.1319  
## age_years     0.9978     1.0023    0.9930    1.0025  
## sourceother   1.1953     0.8366    1.0133    1.4100  
## days_onset_hosp 0.9012     1.1097    0.8764    0.9267  
  
##  
## Concordance= 0.566  (se = 0.009 )  
## Likelihood ratio test= 71.31  on 4 df,  p=1e-14  
## Wald test           = 59.22  on 4 df,  p=4e-12  
## Score (logrank) test = 59.54  on 4 df,  p=4e-12
```

Coefficient

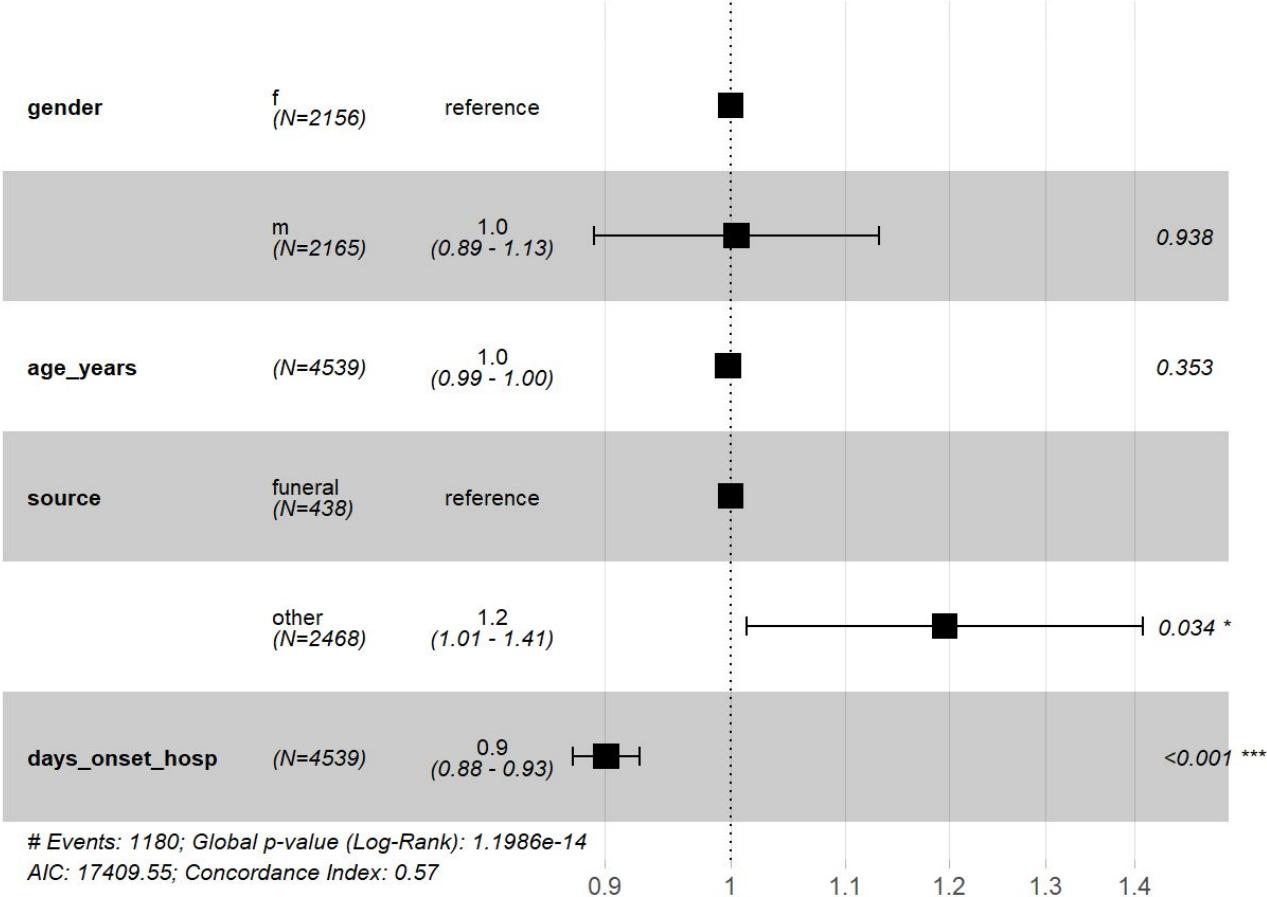
Significant

Hazard

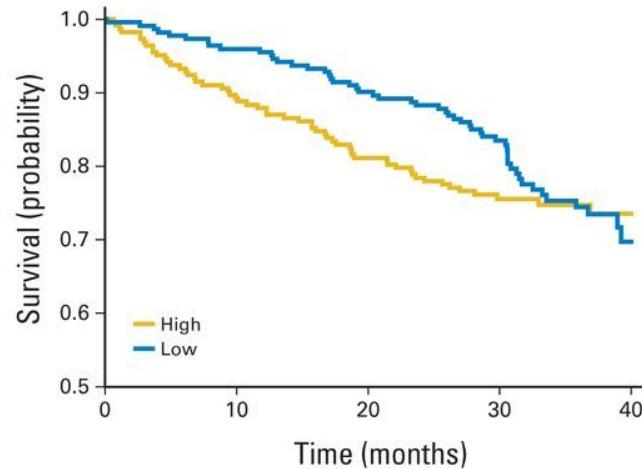
```
## Call:  
## coxph(formula = Surv(futime, event) ~ gender + age_years + source +  
##         days_onset_hosp, data = linelist_surv)  
##  
##   n= 2772, number of events= 1180  
##   (1767 observations deleted due to missingness)  
##  
##             coef exp(coef)  se(coef)      z Pr(>|z|)  
## genderm      0.004710  1.004721  0.060827  0.077  0.9383  
## age_years     -0.002249  0.997753  0.002421 -0.929  0.3528  
## sourceother    0.178393  1.195295  0.084291  2.116  0.0343 *  
## days_onset_hosp -0.104063  0.901169  0.014245 -7.305 2.77e-13 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.  
##  
##             exp(coef) exp(-coef) lower .95 upper .95  
## genderm      1.0047     0.9953    0.8918    1.1319  
## age_years     0.9978     1.0023    0.9930    1.0025  
## sourceother    1.1953     0.8366    1.0133    1.4100  
## days_onset_hosp 0.9012     1.1097    0.8764    0.9267  
##  
## Concordance= 0.566  (se = 0.009 )  
## Likelihood ratio test= 71.31  on 4 df,  p=1e-14  
## Wald test          = 59.22  on 4 df,  p=4e-12  
## Score (logrank) test = 59.54  on 4 df,  p=4e-12
```

Confident Intervals

Hazard ratio



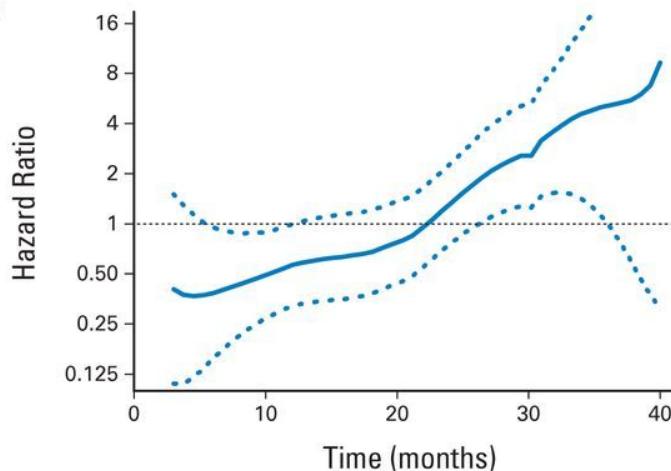
Others

A

No. at risk	
High	223
Low	222

199 179 122 20

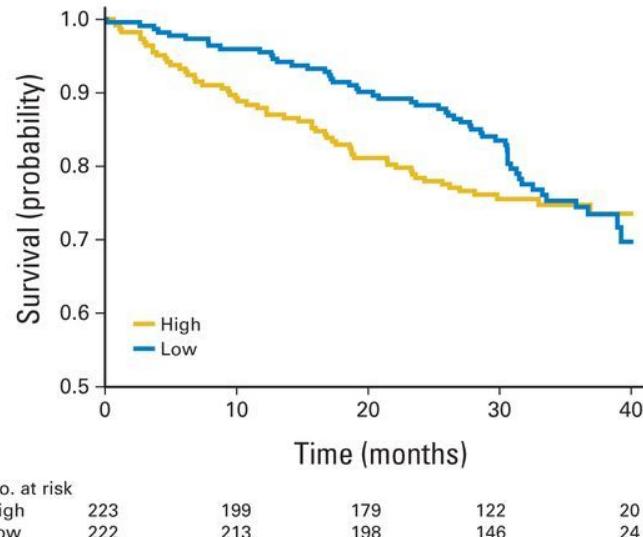
213 198 146 24

B

STATISTICS IN ONCOLOGY 2014

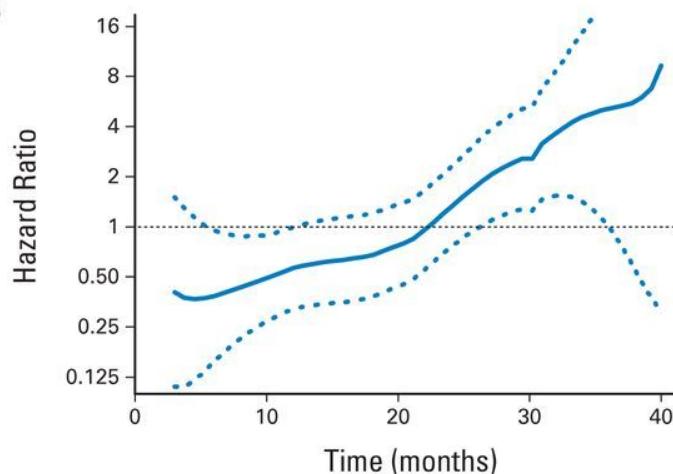
Moving Beyond the Hazard Ratio in Quantifying the
Between-Group Difference in Survival Analysis

DOI: 10.1200/JCO.2014.55.220

A

Cox's proportional hazard model

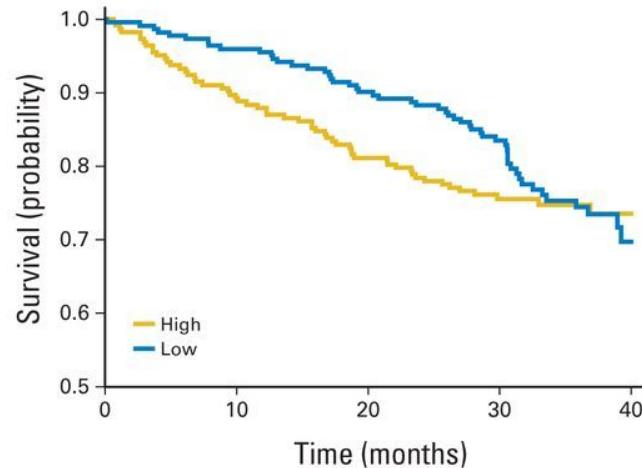
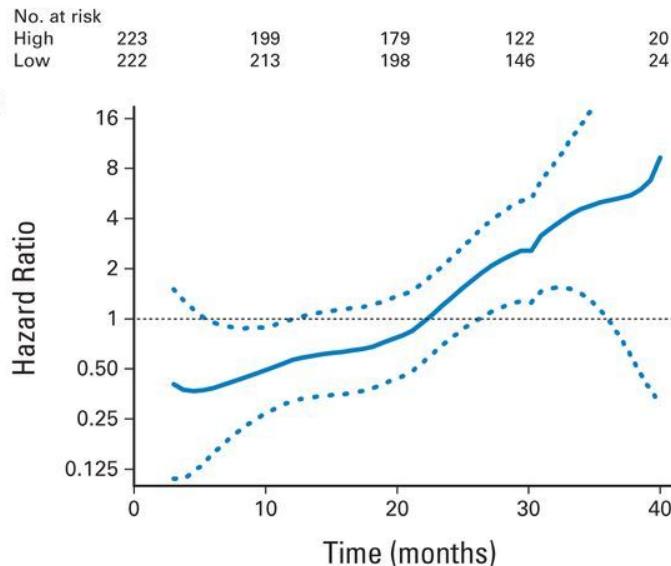
0.87 [0.60 , 1.27] n.s.

B

STATISTICS IN ONCOLOGY 2014

Moving Beyond the Hazard Ratio in Quantifying the Between-Group Difference in Survival Analysis

DOI: 10.1200/JCO.2014.55.220

A**B**

Cox's proportional hazard model

0.87 [0.60 , 1.27] n.s.

Ratio of Restricted Mean Survival Times

0.68 [0.47 , 0.98] *

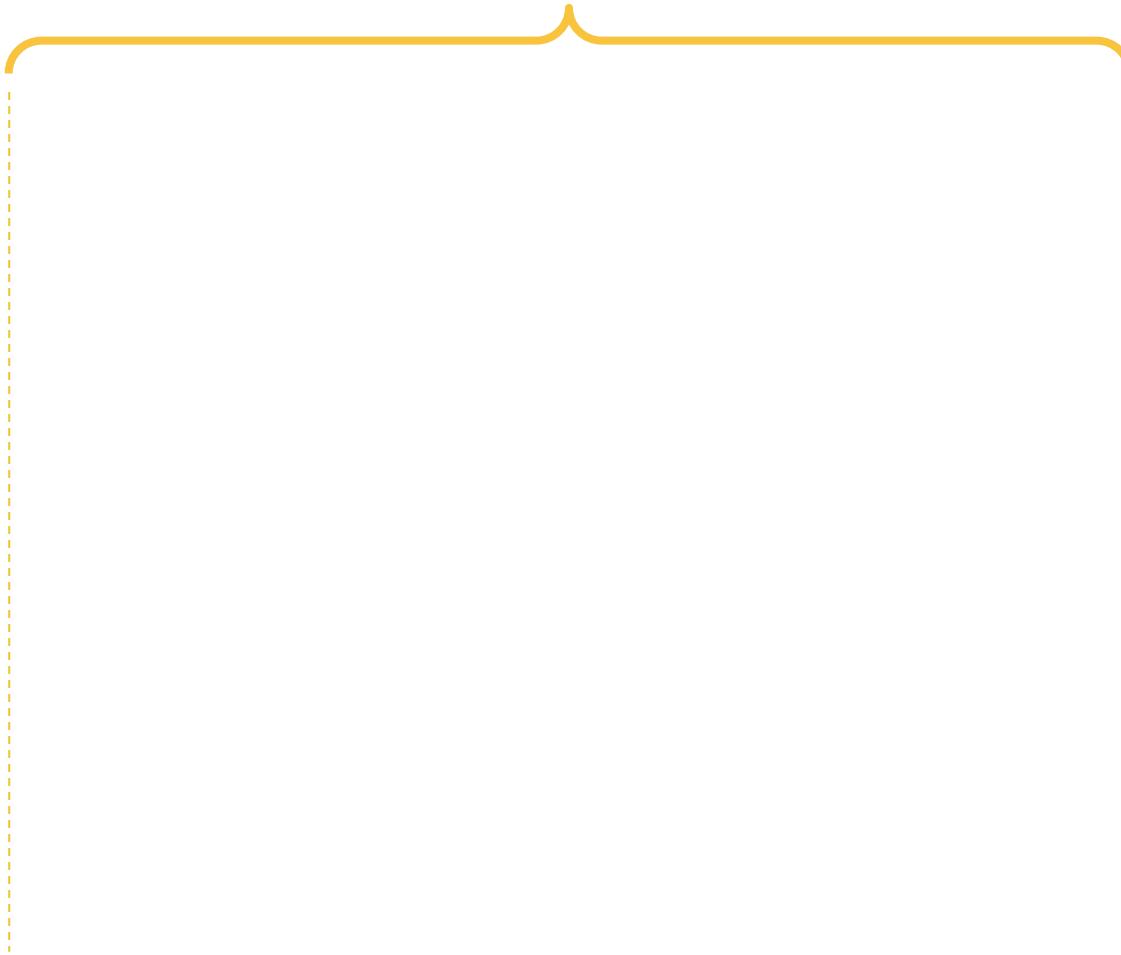
STATISTICS IN ONCOLOGY 2014

Moving Beyond the Hazard Ratio in Quantifying the Between-Group Difference in Survival Analysis

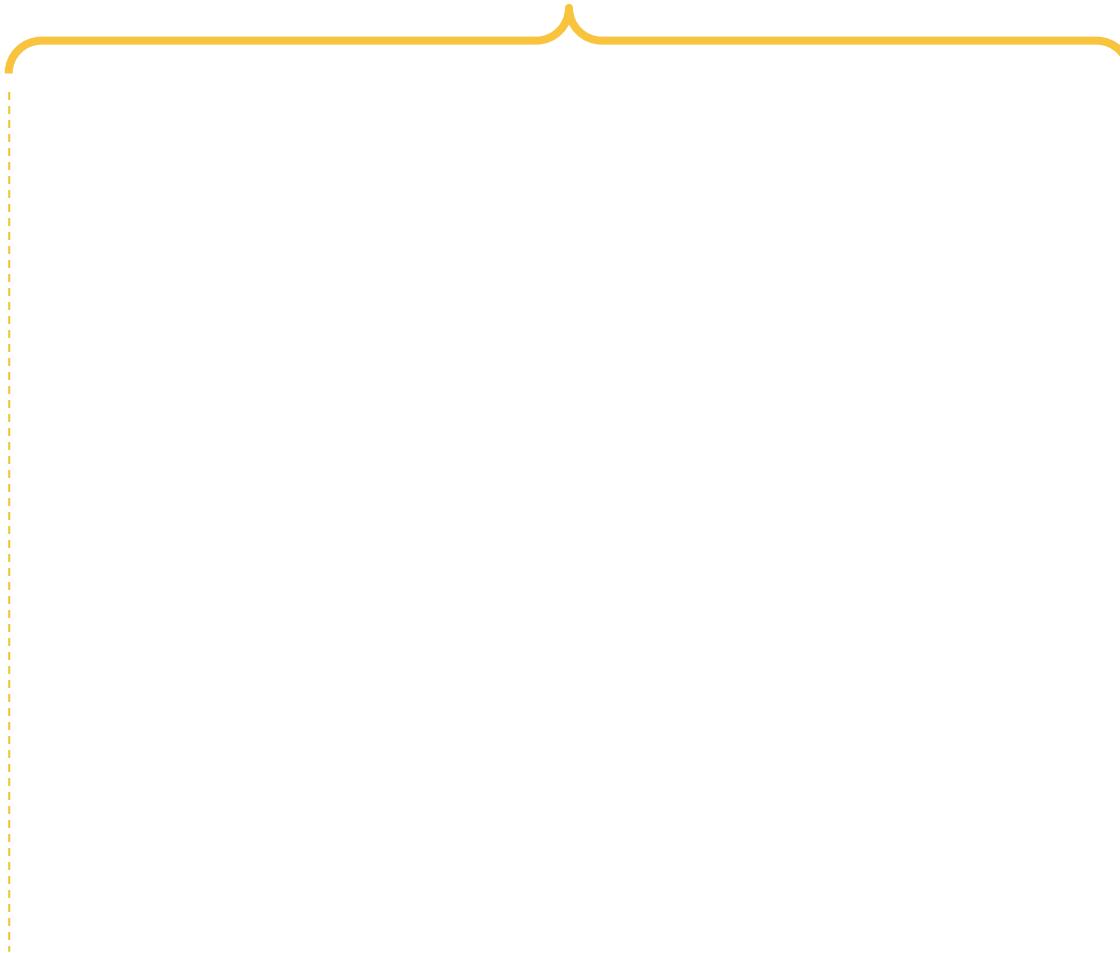
DOI: 10.1200/JCO.2014.55.220

5 - CENSORED DATA

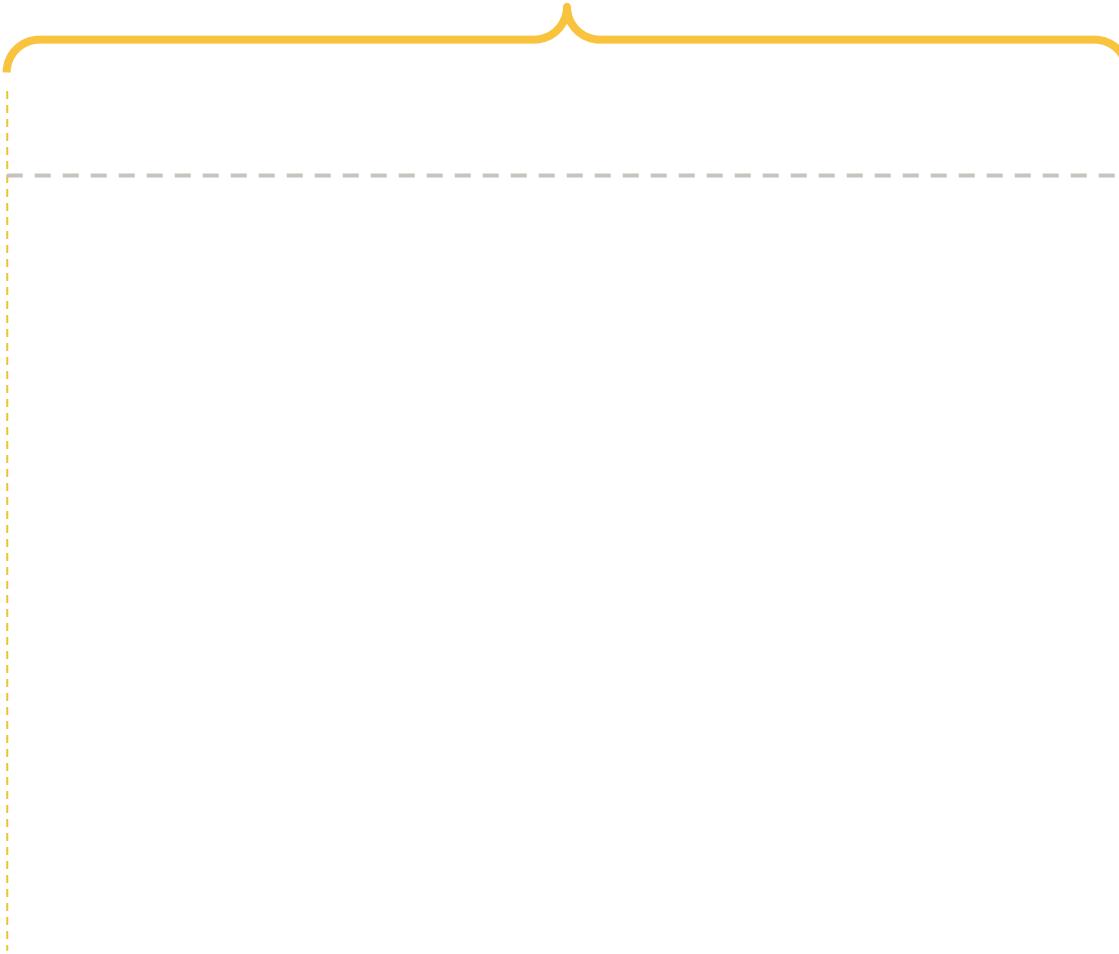
Time of the study



Time of the study



Time of the study



Time of the study



Right



Time of the study



Right

Right

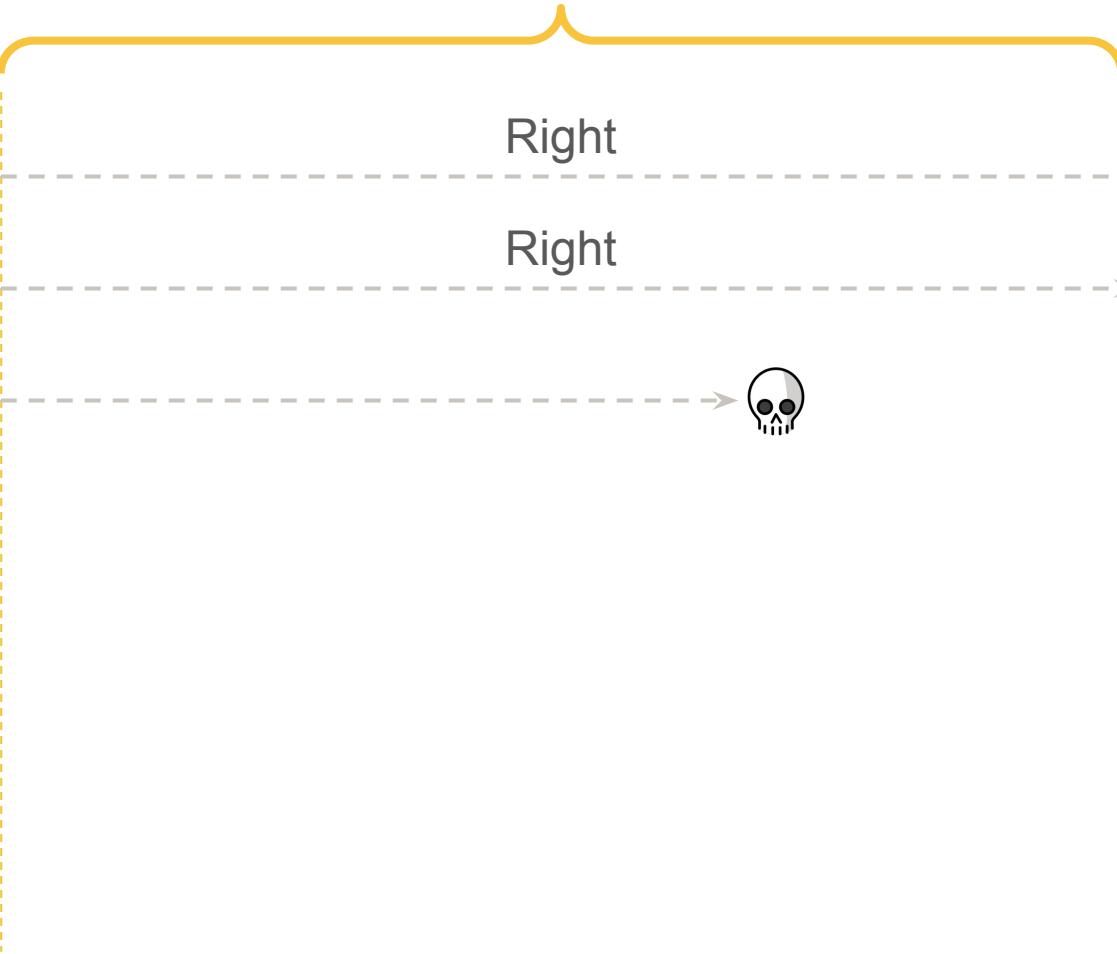


Time of the study



Right

Right



Time of the study



Right



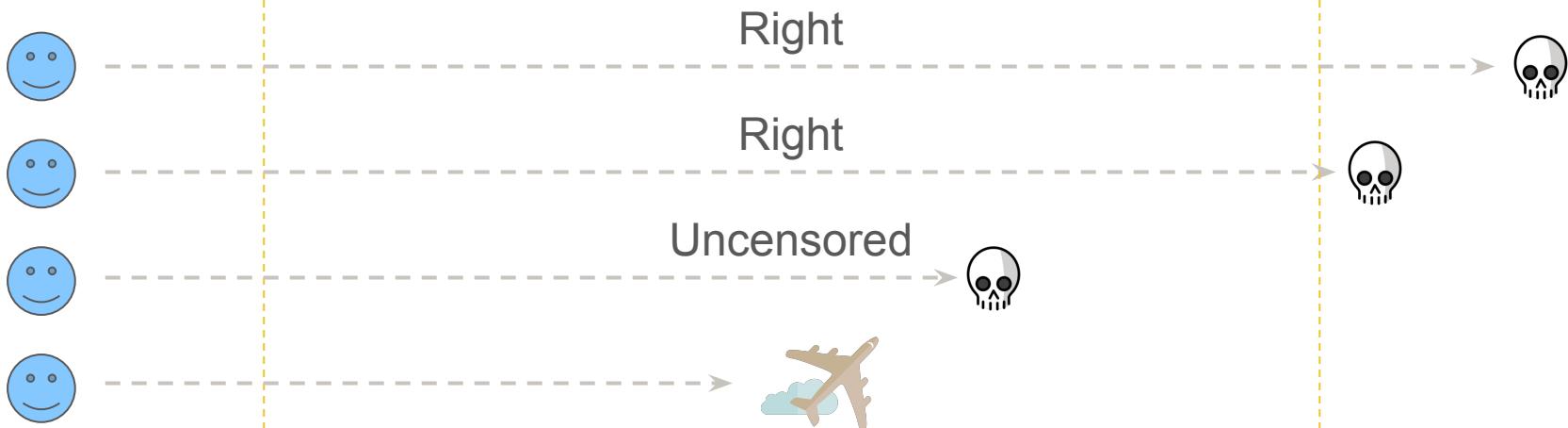
Right



Uncensored



Time of the study



Time of the study



Right



Right



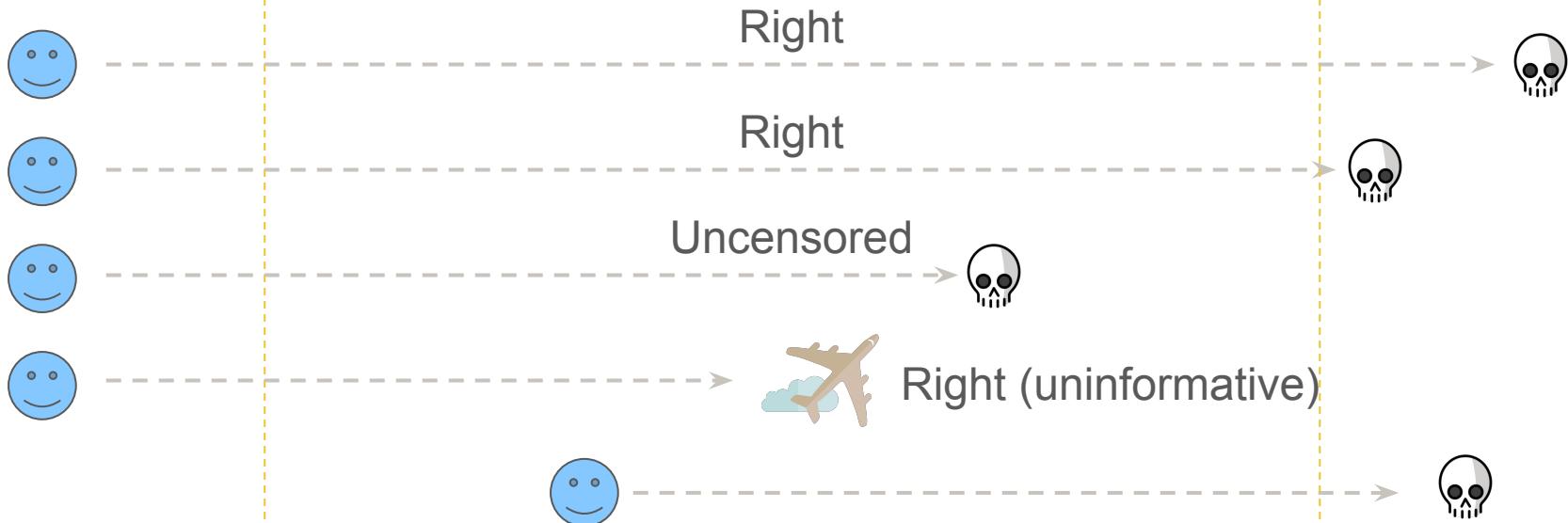
Uncensored



Right (uninformative)



Time of the study



Time of the study



Right



Right



Uncensored



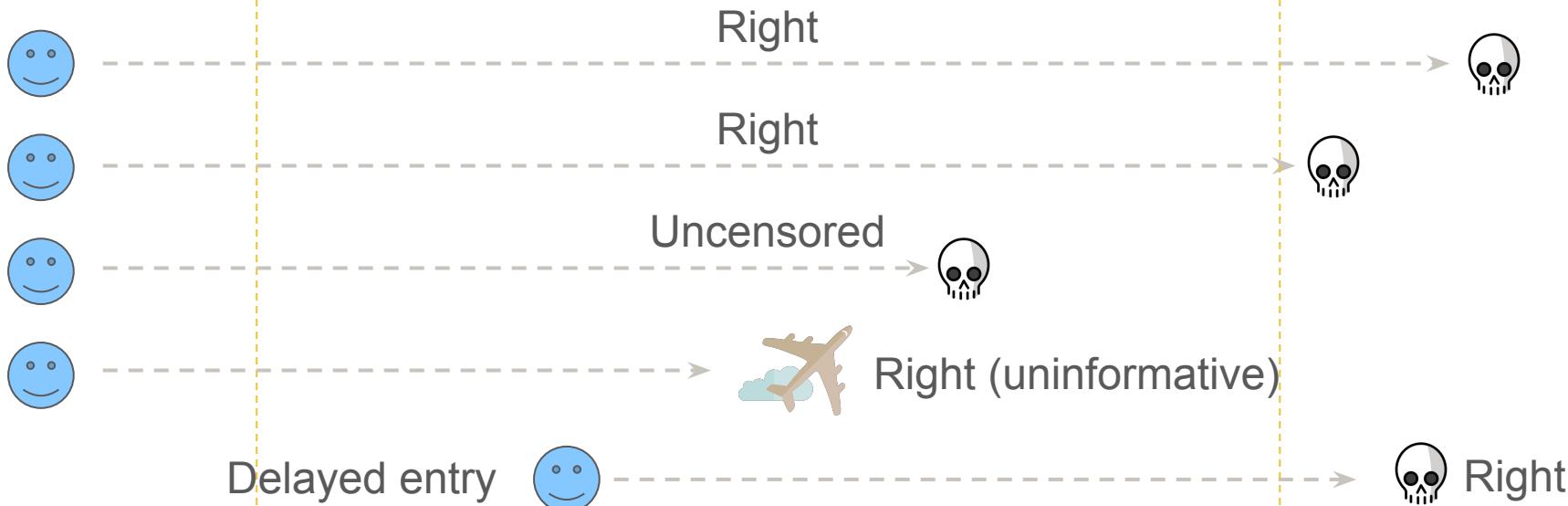
Right (uninformative)

Delayed entry



Right

Time of the study

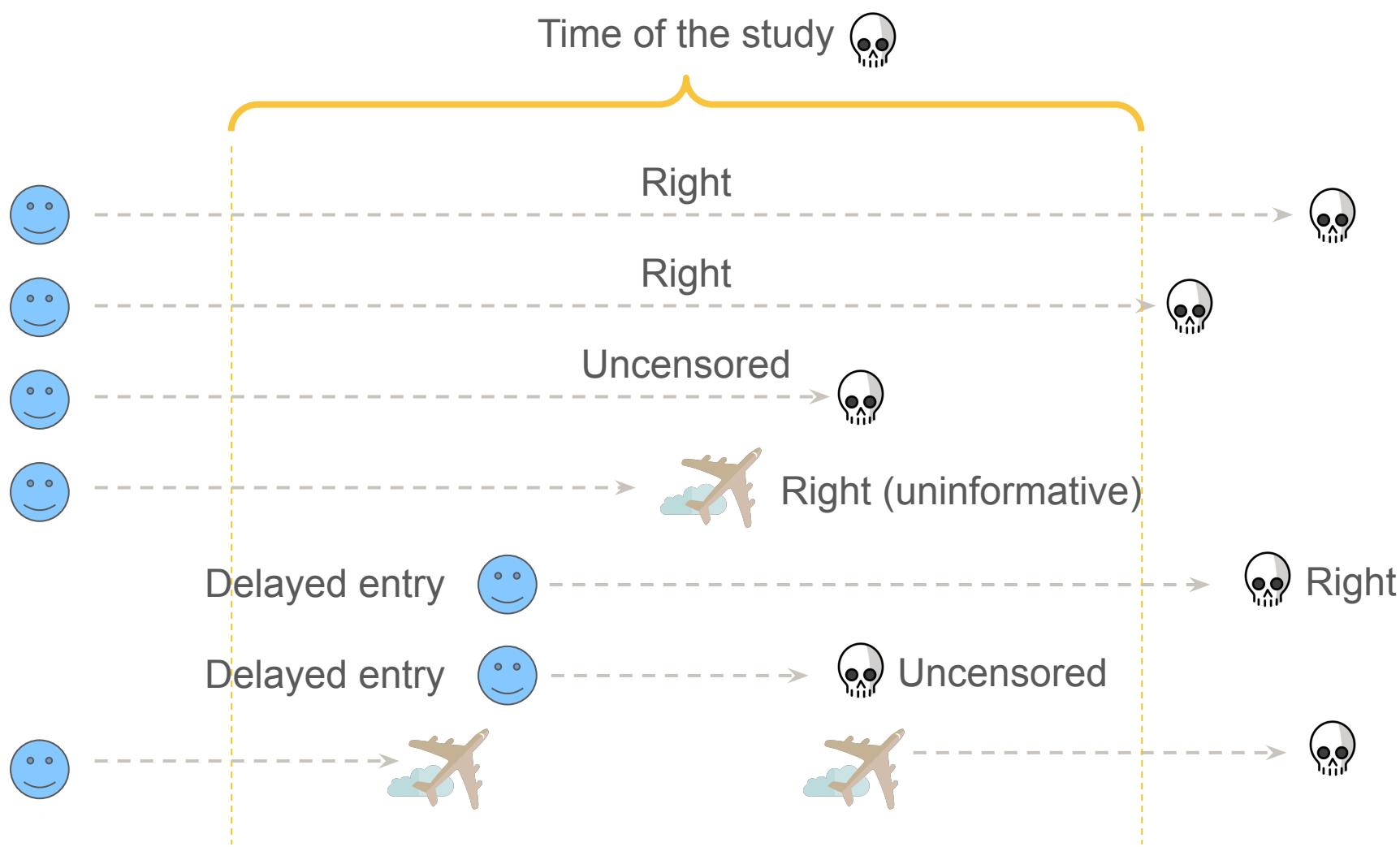


Right

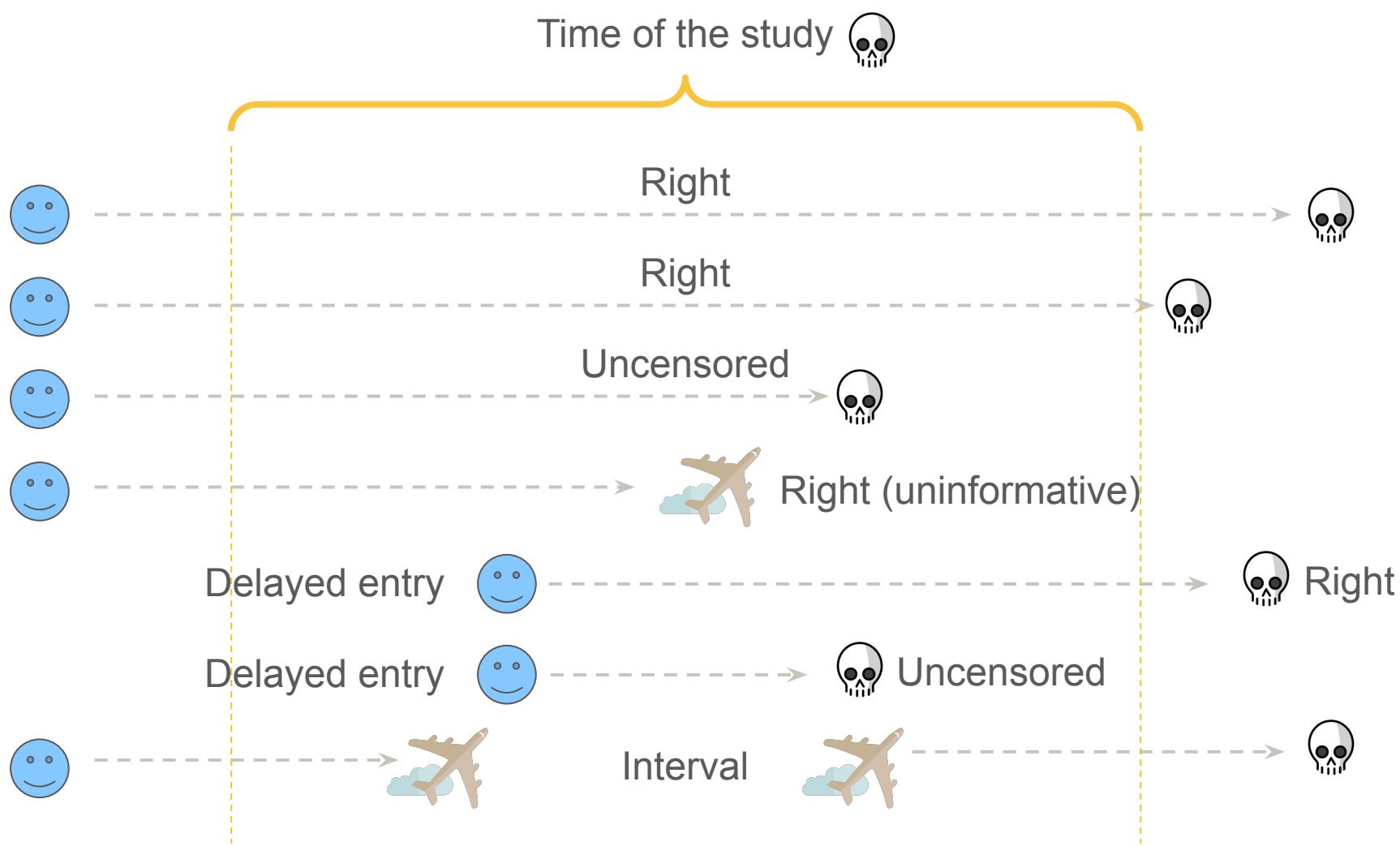
Right

Uncensored

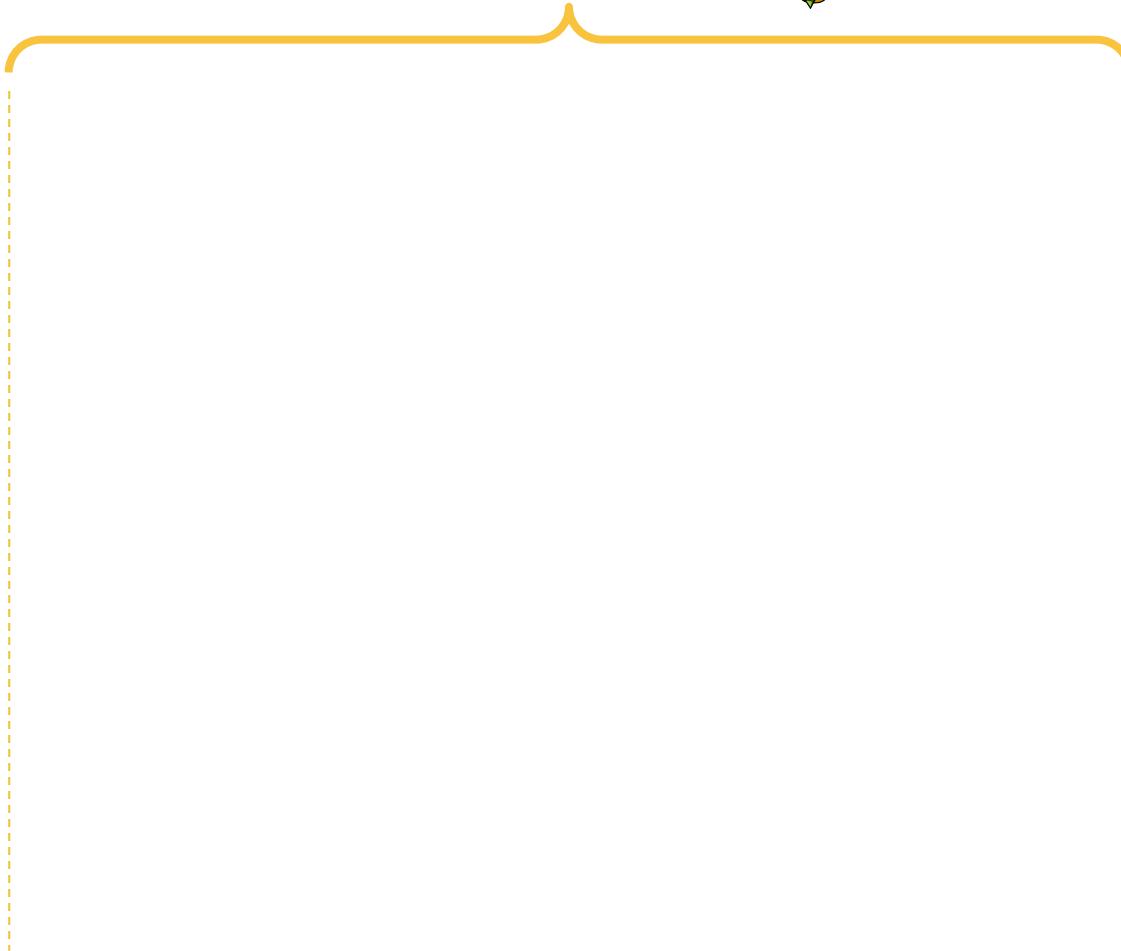
Time of the study



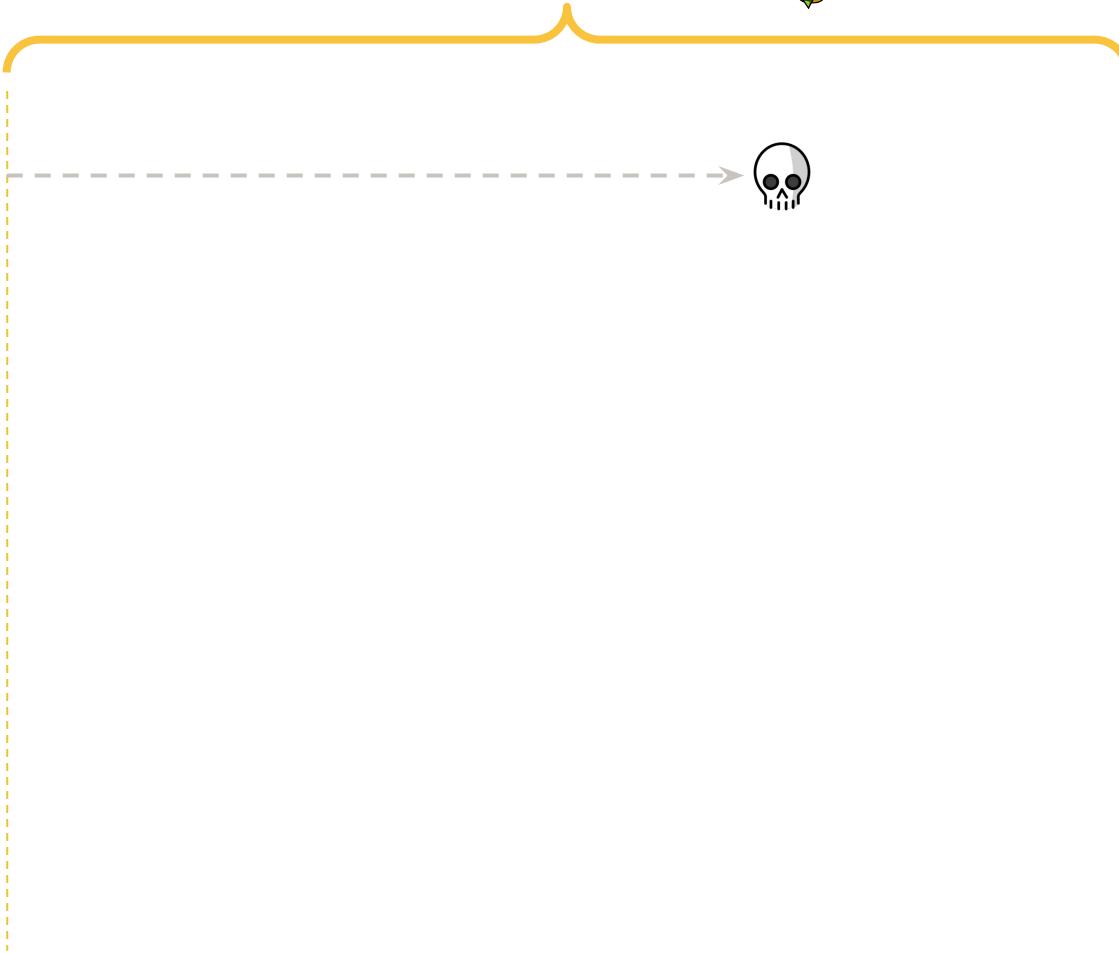
Time of the study



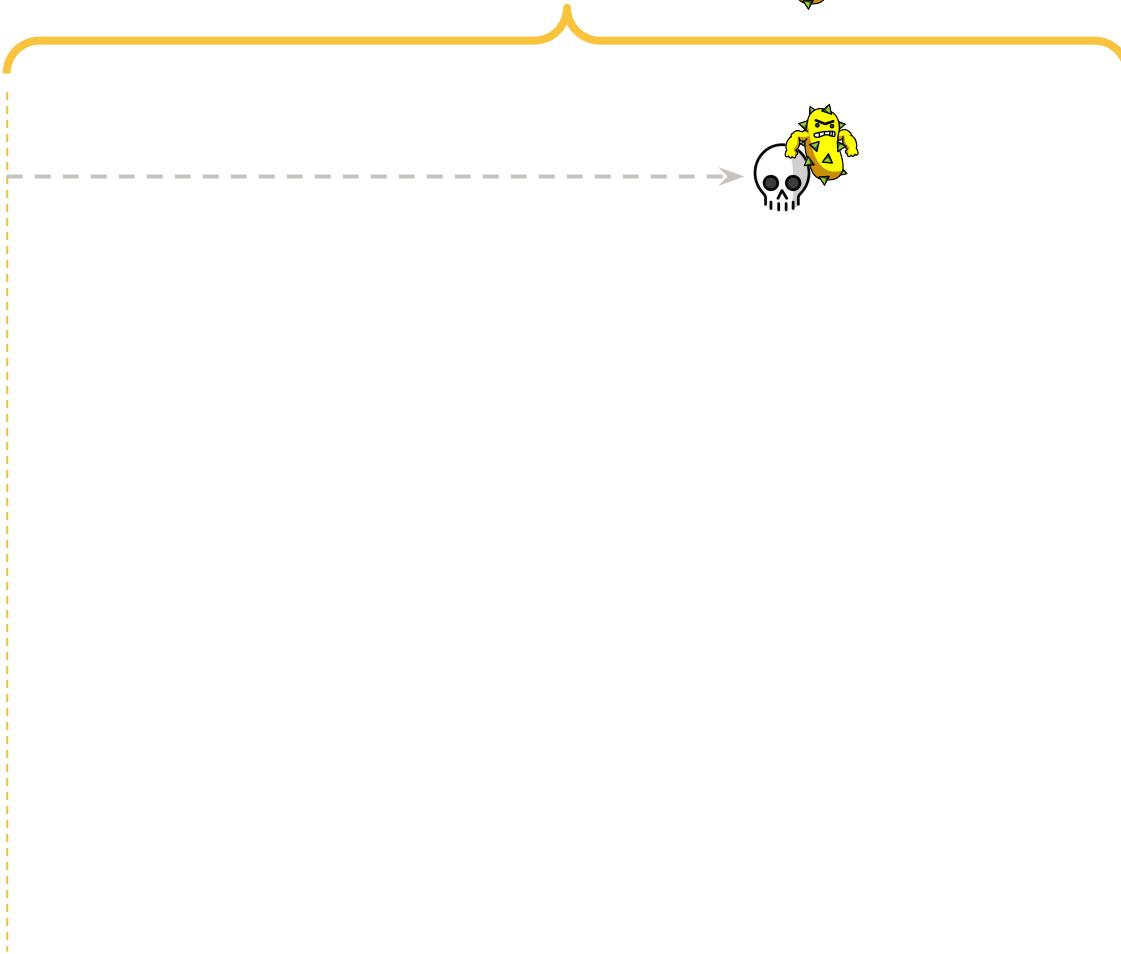
Time of the study



Time of the study



Time of the study



Time of the study



Uncensored



Time of the study



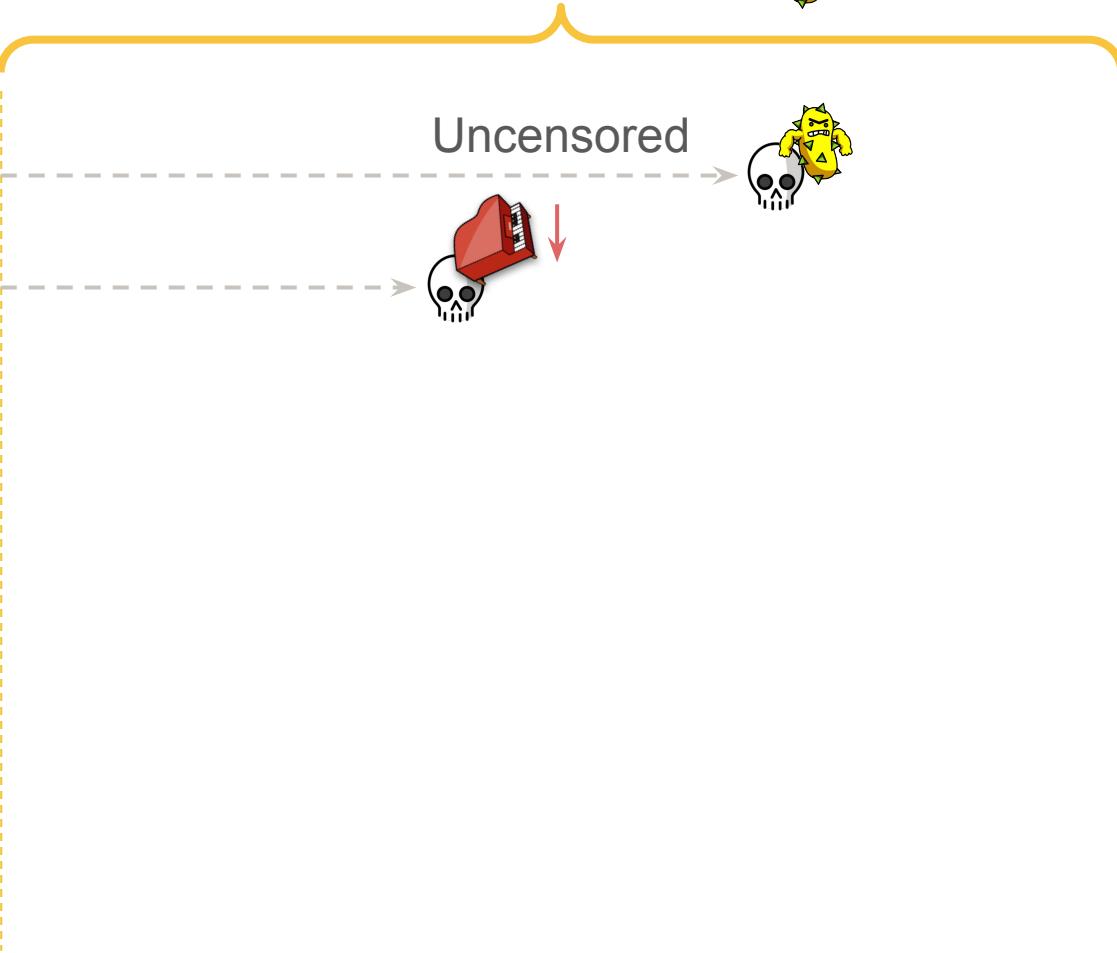
Uncensored



Time of the study



Uncensored



Time of the study



Uncensored



Informative

Time of the study



Uncensored

Informative



Time of the study



Uncensored



Informative



nature reviews clinical oncology

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [nature reviews clinical oncology](#) > [comment](#) > [article](#)

Comment | [Published: 09 April 2020](#)

Informative censoring — a neglected cause of bias in oncology trials

[Arnoud J. Templeton](#) [orcid.org/0000-0002-0847-7087](#)^{1,2}, [Eitan Amir](#) [orcid.org/0000-0002-3706-525X](#)³ &
[Ian F. Tannock](#) [✉³](#)

Time of the study



Informative

Uncensored



Time of the study



Informative

Uncensored



Type II Censoring

Time of the study



Informative

Uncensored



Type II Censoring



Time of the study



Time of the study



Time of the study

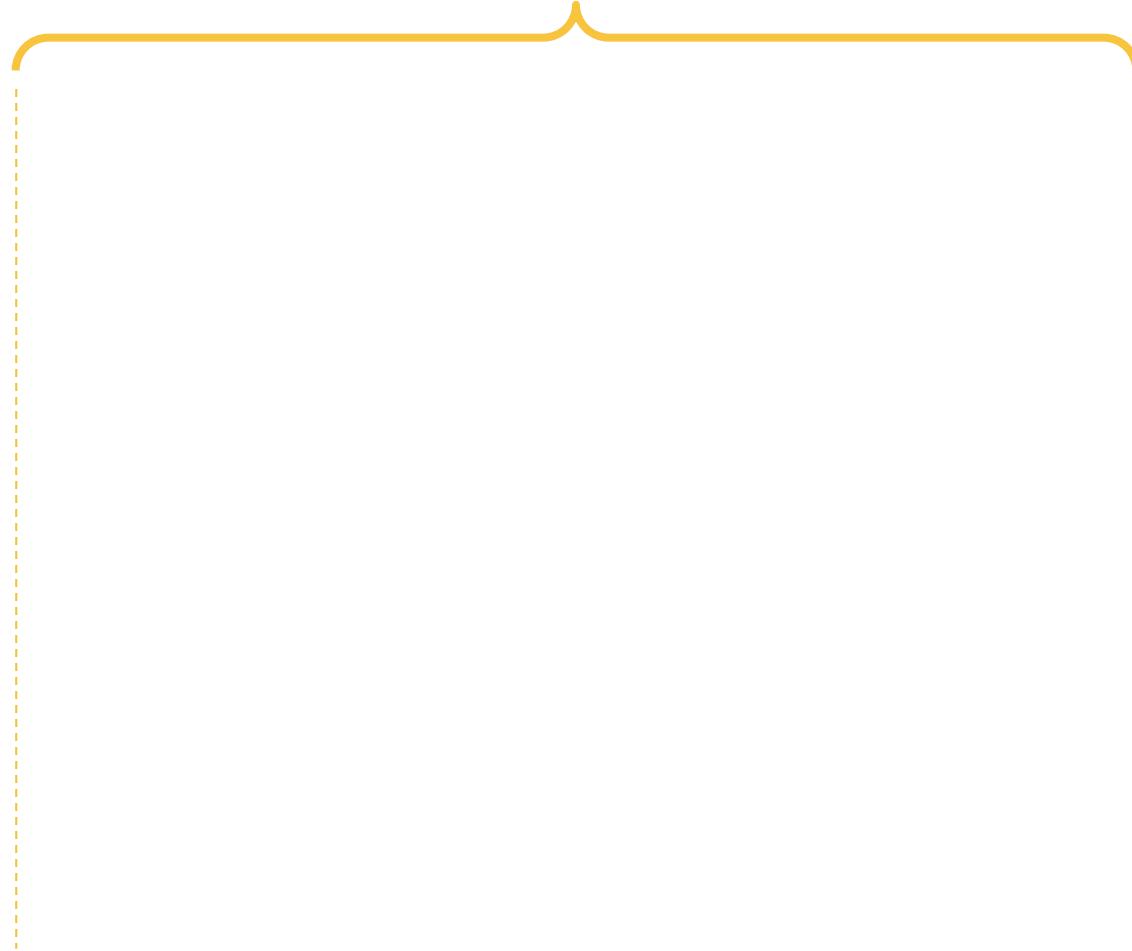


Type I Censoring

Time of the study

1700

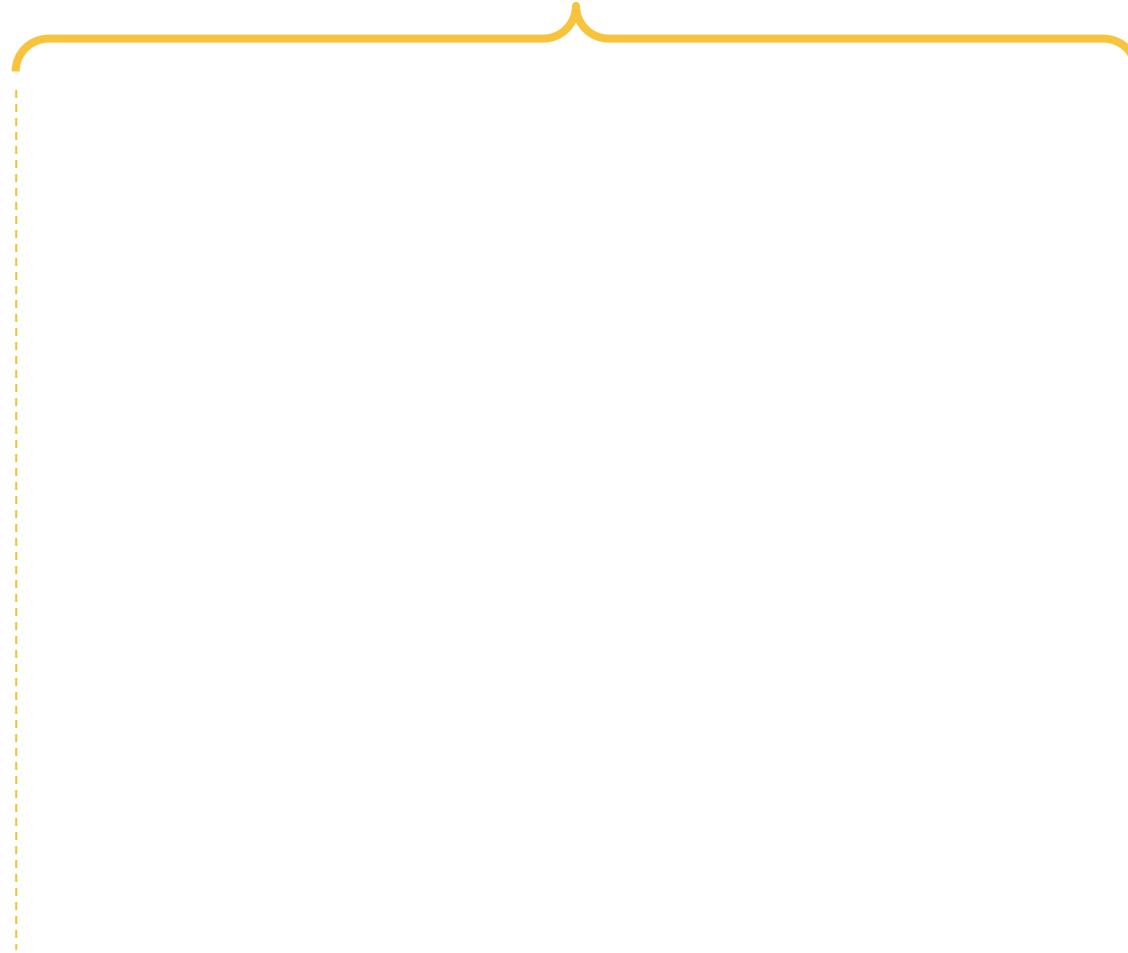
1800

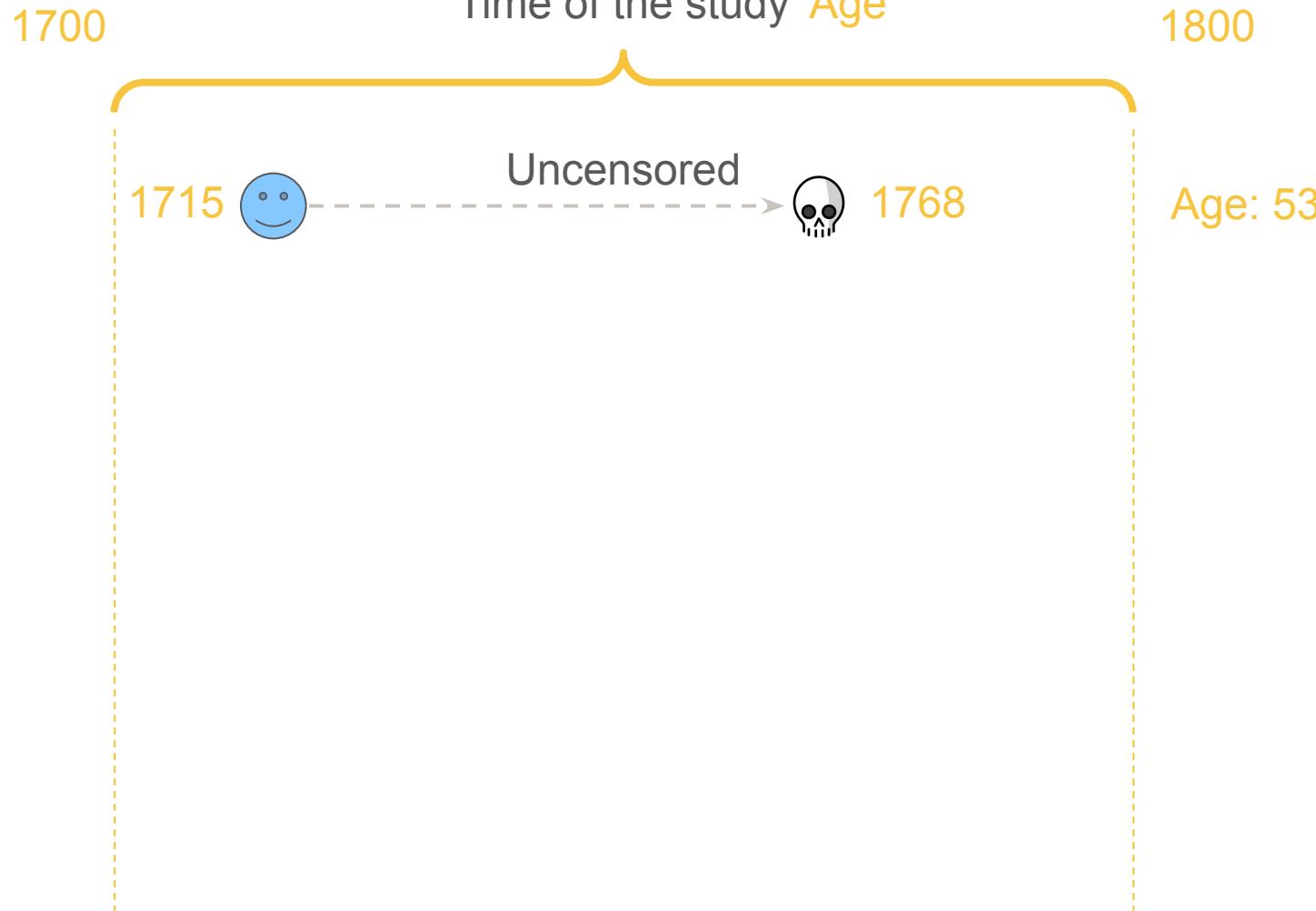


Time of the study **Age**

1700

1800





1700 Time of the study Age 1800

1715 

Uncensored



1768

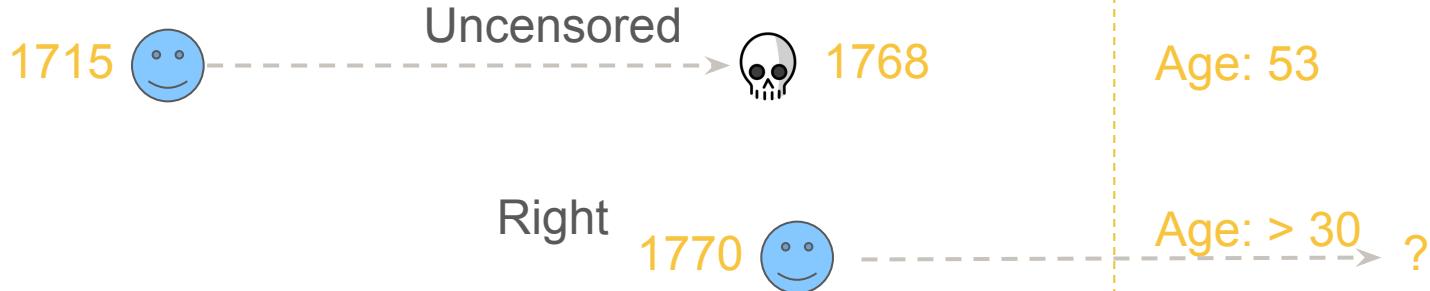
Age: 53

Right

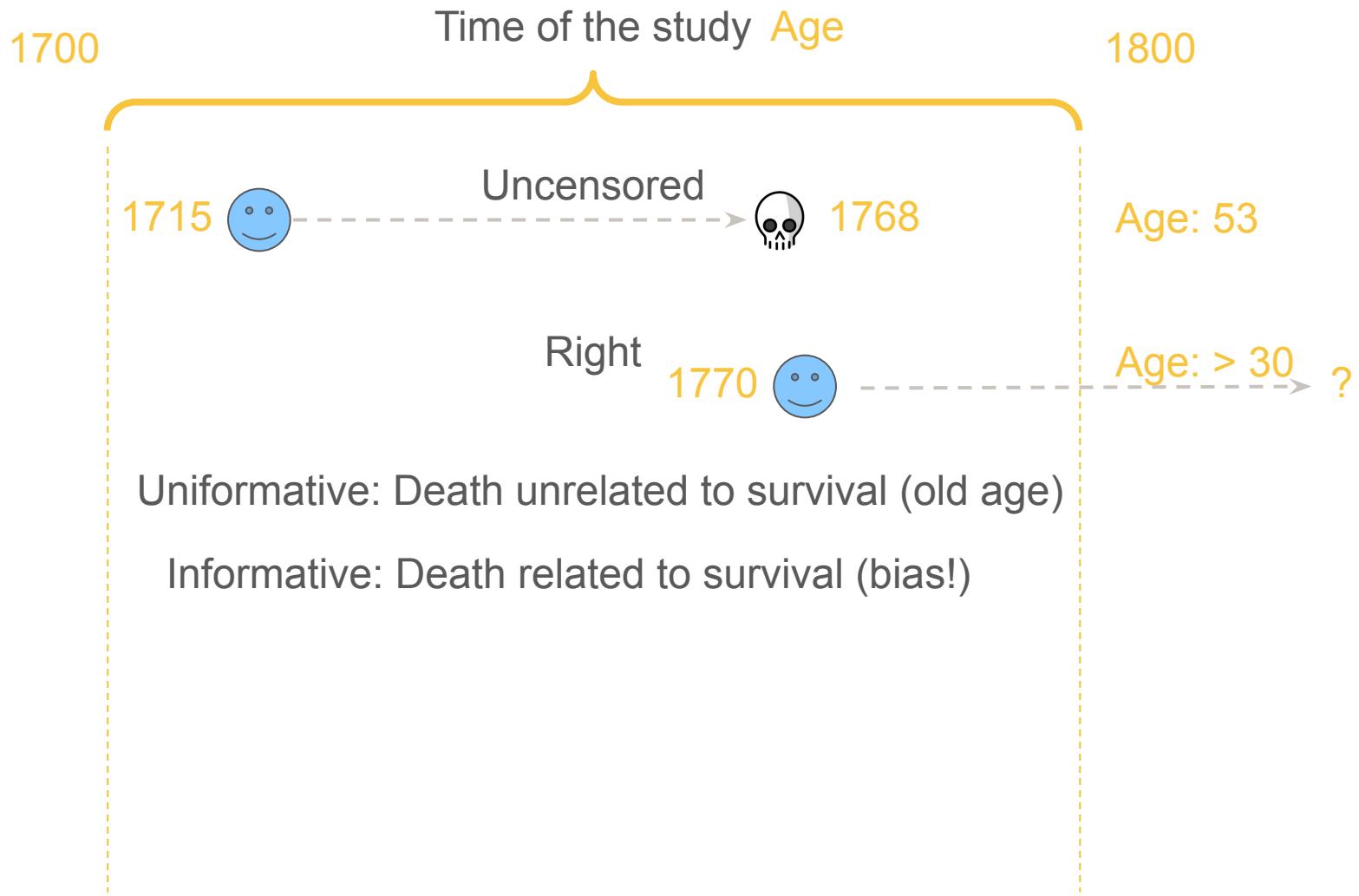
1770 

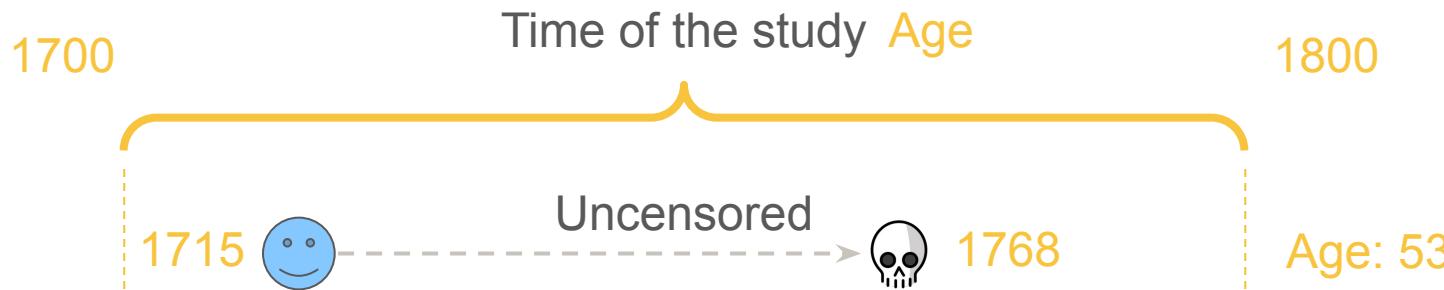
Age: > 30 ?

1700 Time of the study Age 1800



Uniformative: Death unrelated to survival (old age)

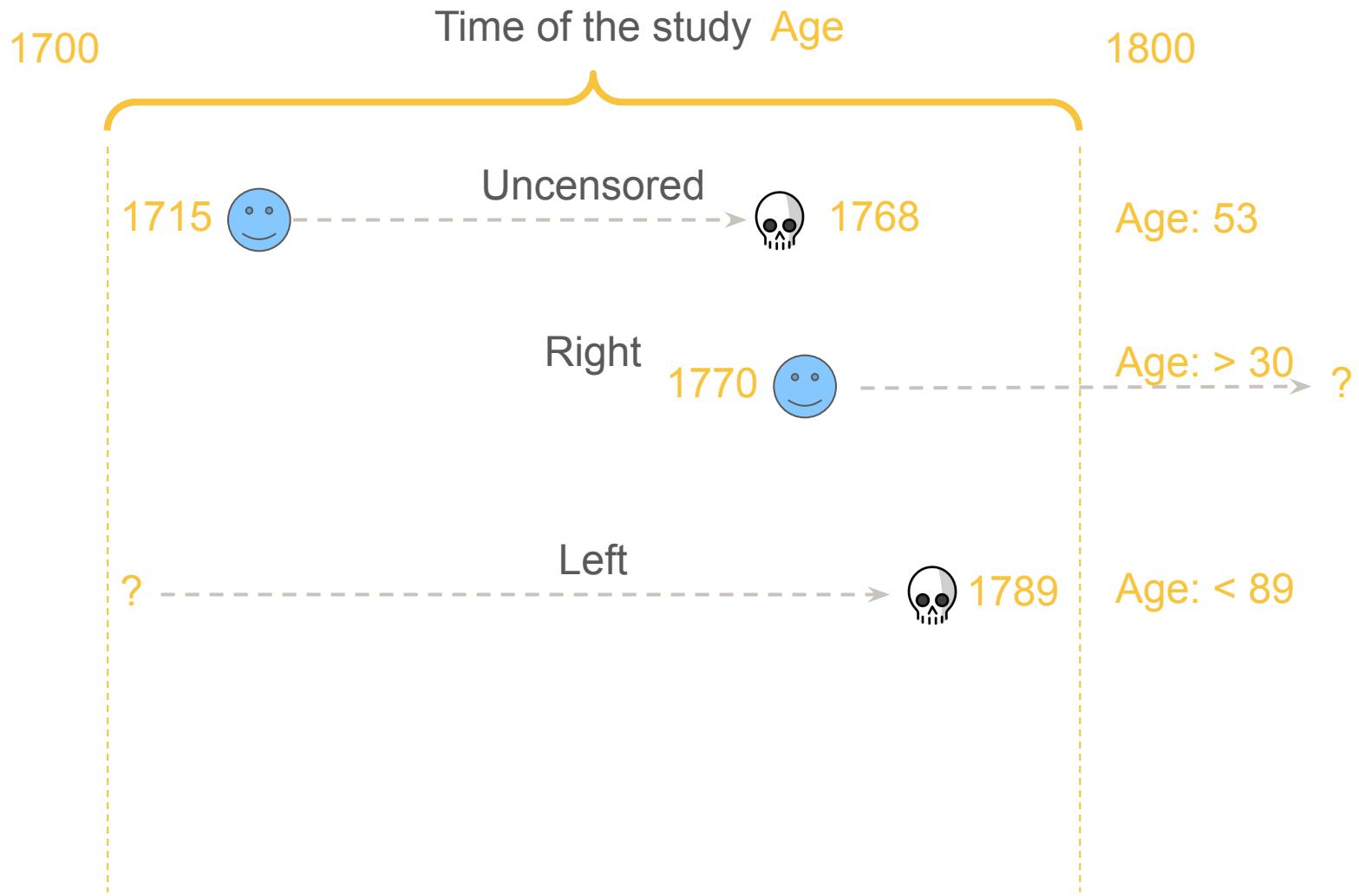




Uniformative: Death unrelated to survival (old age)

Informative: Death related to survival (bias!)

	1772	Bubonic Plague	50.000
	1773	Bubonic Plague	2.000.000
	1775	Smallpox	15.000
(Au)	1789	Smallpox	150.000
	1800	Yellow Fever	60.000



Presentation

