INWS0038 Longitudinal and Multilevel Modelling II - Event History Analysis

22.04 – 10.05 2024 (3 ECTS)

Linus Andersson[1,2]

[1]*University of Turku,* [2]*Stockholm University*

linus.andersson@utu.fi

INWS0038 Longitudinal and Multilevel Modelling II - Event History Analysis

lecture III – Parametric models

Monday 29.04

Linus Andersson[1,2]

[1]*University of Turku,* [2]*Stockholm University*

linus.andersson@utu.fi

Practice does not make perfect,
practice makes permanent

- Anonymous

Parametric models I

- We covered:
  - Concepts of EHA
  - How to calculate and show S(t) and h(t)
  - Using Life tables & Kaplan-Meier
  - That is, ”Descriptive”, non-parametric EHA

- Now, parametric models
  - Regression framework
  - Assumptions about S(t) and h(t)
  - Efficient, but less robust
  - Tests theory about time process
  - New estimator, new interpretation of the estimate
  - Many choices to be made

Scheduling

Activity	Lesson topic	Keywords	Homework
Lesson 1.	Introduction Key Concepts	Process time, Cencoring, Time-to-Event, Continous and discrete time	
Lesson 2.	Key estimates Descriptive models	Kaplan-meyer, Density, Cum. Distribution function, Survival and Hazard funciton, Kaplan-Meyer, Life tables	✓
Lesson 3.	Key Concepts and estimates for Parametric models	Exponential and Piece-wise exponential models, Shape parameter, the proportional hazard assumption, hazard ratios	✓
Lab 1. Non-Parametric models			
Lesson 4	Discrete and Continous models Data structure	Time-varying variables, Cox, Logit	✓
Lab 2. Parametric models I.			
Lesson 5.	Piecing in together + Extensions	Competing risk models, Causality, heterogeneity	
Lab 3. Parametric models II.			
Lesson 6.	Discussion. Presentations.		Presentations

Parametric models I

	Non-Parametric	Parametric	The parametric option imply:
Tools	Life table or Kaplan-Meier	Numerous	Model choice
Metrics	S(t), h(t)	Hazard ratio, log-hazard, AFT	New inter-pretations
Assumptions about T & h(t)	no	Specific shapes	Describes a model, not data

Key estimates in EHA

1. Model assumptions about the shape of the hazard
2. The proportional hazards assumption
3. Interpretation of estimates

Parametric models I – model choice

- Primarily, model choice is based on data and the phenomena of inquiry
  - The shape of the time process (constant, increasing, decreasing, etc., over time)
  - The quality of the time process (continuous or discrete time events)
  - Efficiency versus robustness (Data size, nr of assumptions, degrees of freedom)

Parametric models I – model choice – Distribution of T

Some very common models

- Assumptions about T = the distribution of events over time
- and its association with the hazard estimate

Model	Assumed shape of T
Cox	
Exponential	
Piece-wise constant exponential	
Gompertz	
Log-normal, Log-logistic	

RECAP: Functions

Survivor function S(t)

Probability of not having had an event up to time t

F(t-1)

Cumulative distribution function F(t)

Probability of ever having had the event up to time t

1-S(t)

Probability density function f(t)

Nr of event at any given time t

Integral of F(t)

Hazard function h(t)

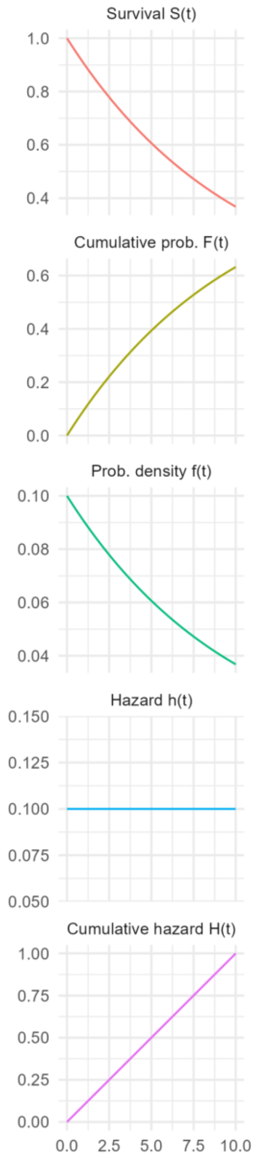
Having a transition at time t, conditional on survival

F(t)/S(t)

Cumulative hazard function h(t)

Sum of hazards. Expected nr of events until t

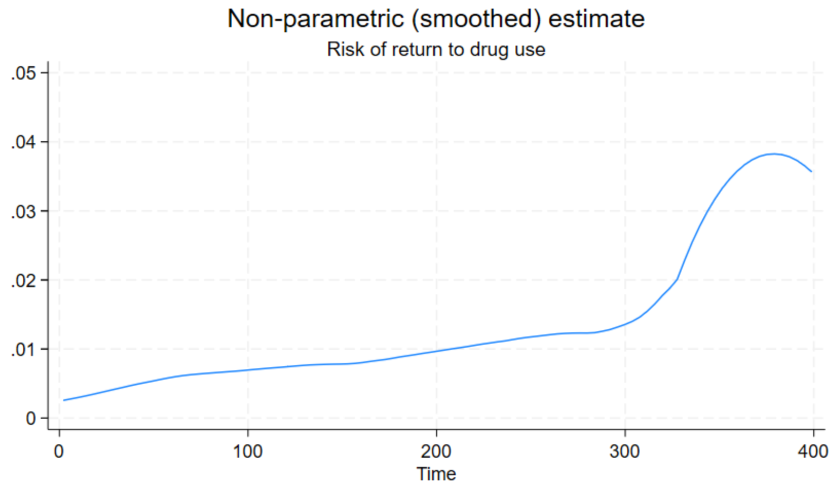
Integral of h(t)



Parametric models I – model choice – Distribution of T

First of, non-parametric smoothed estimate of the hazard

- Useful to compare models against data
- Imposes no assumptions about T



Parametric models I – model choice – Distribution of T

The Cox proptional hazard model

Model	Assumed shape of T
Cox	None
Exponential	
Piece-wise constant exponential	
Gompertz	
Log-logistic	

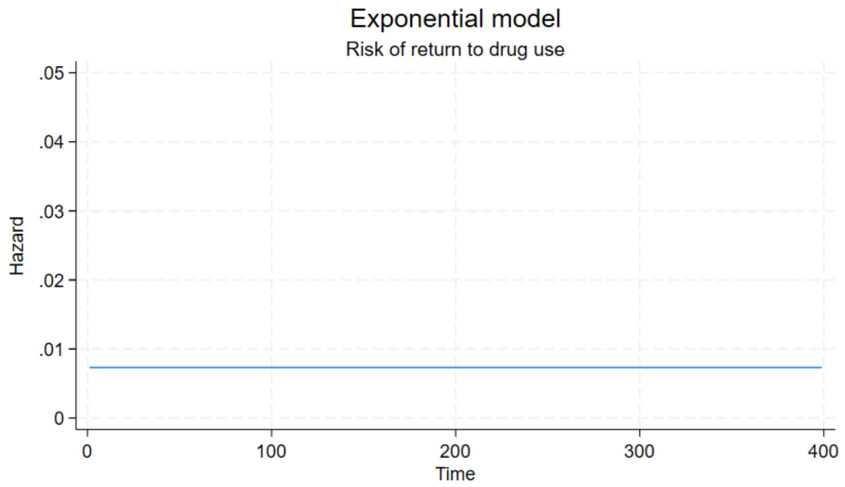
- Imposes no assumptions about T
- Hazard rate vary over time
- No baseline hazard
- More on Cox later

Parametric models I – model choice – Distribution of T

The Exponential model

Model	Assumed shape of T
Cox	
Exponential	Constant
Piece-wise constant exponential	
Gompertz	
Log-logistic	

- Assumes exponential decay of events over time
- Thus assuming a constant hazard rate
- Ln $h(t|\boldsymbol{x})$ = ln $\beta 0 + \boldsymbol{x\beta}$
- Baseline (e.g., the constant) is flat
- Same risk of event no matter time (“memory-less”)

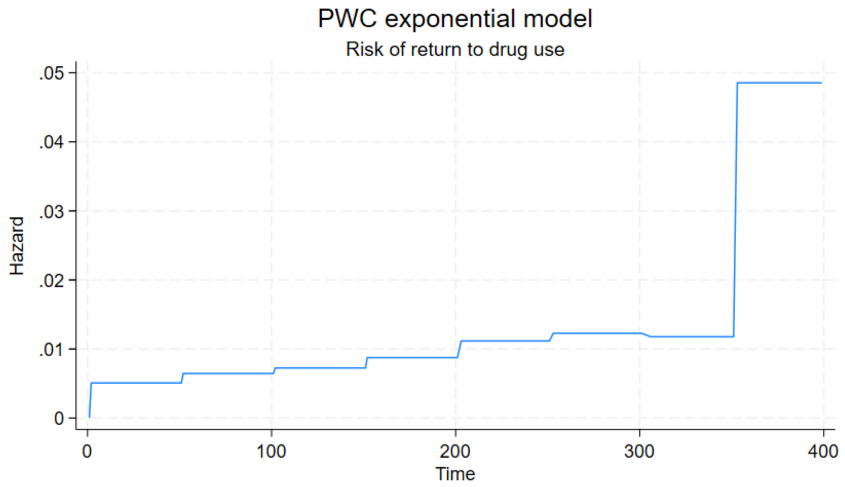


Parametric models I – model choice – Distribution of T

The Piece-wise Constant Exponential model

Model	Assumed shape of T
Cox	
Exponential	
Piece-wise constant exponential	Any shape across intervals
Gompertz	
Log-logistic	

- Extension of the constant exponential model
- Divides time into intervals
- Hazard rate constant within but vary between intervals
- $\ln h(t|\boldsymbol{x}) = \ln(\beta 01 p1 + \cdots + \beta 01{+}pk) + \boldsymbol{x\beta}$

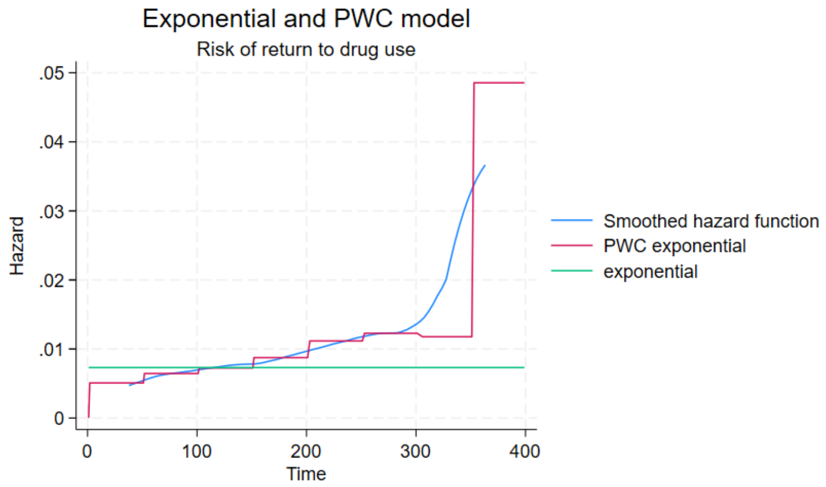


Parametric models I – model choice – Distribution of T

Summary so far:

Model	Assumed shape of T
Cox	
Exponential	Constant
Piece-wise constant exponential	Any shape across intervals
Gompertz	
Log-logistic	

- Exponential or PCW exponential model?
  - Which is closest to raw data?
  - PWC!
  - But remember – PWC requires sufficient events and person-time *at each interval (N and events required) – not efficient!*

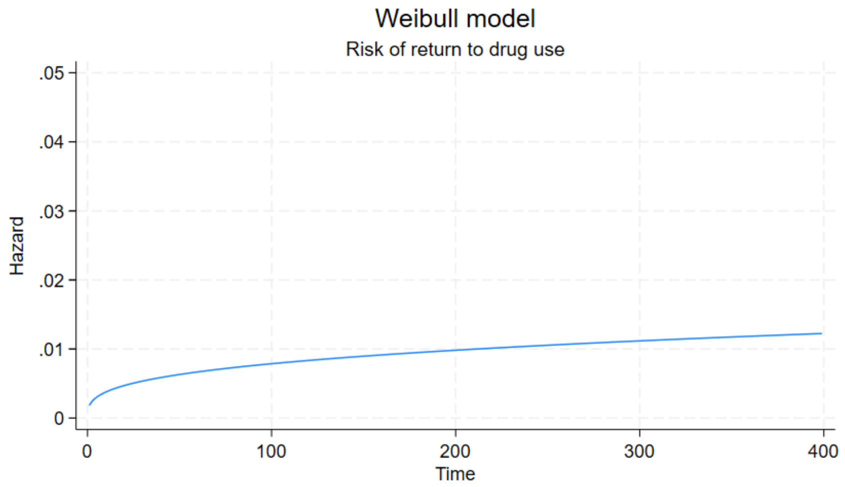


Parametric models I – model choice – Distribution of T

The Weibull model

Model	Assumed shape of T
Cox	
Exponential	
Piece-wise constant exponential	
Weibull	Increasing or decrease
Log-logistic	

- Hazard rate can increase, decrease or be constant
- In addition to shape, has a scale parameter
- Very versatile

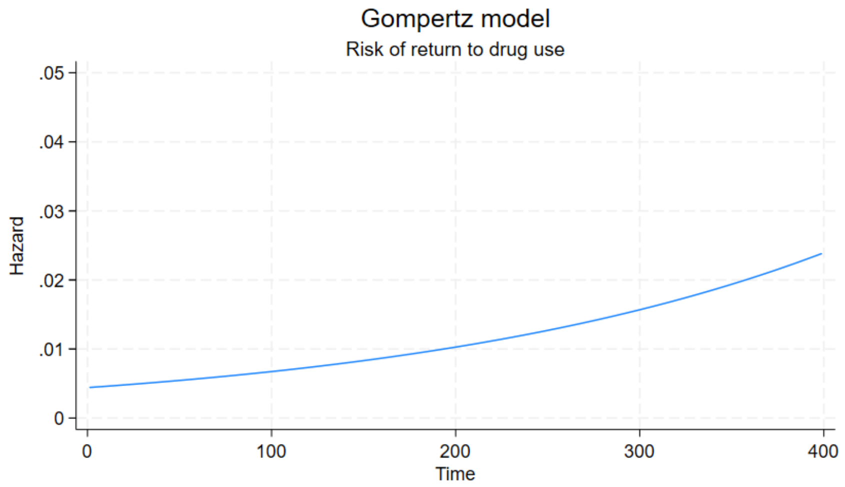


Parametric models I – model choice – Distribution of T

The Gompertz model

Model	Assumed shape of T
Cox	
Exponential	
Piece-wise constant exponential	
Gompertz	Increasingly increasing
Log-logistic	

- Assumes a Gompertz distribution
- Thus, hazard increases exponentially with time
- As Weibull both shape, and scale parameter, but also e = intensity of increase
- Useful when risk of event accelerates with time
- Such as death

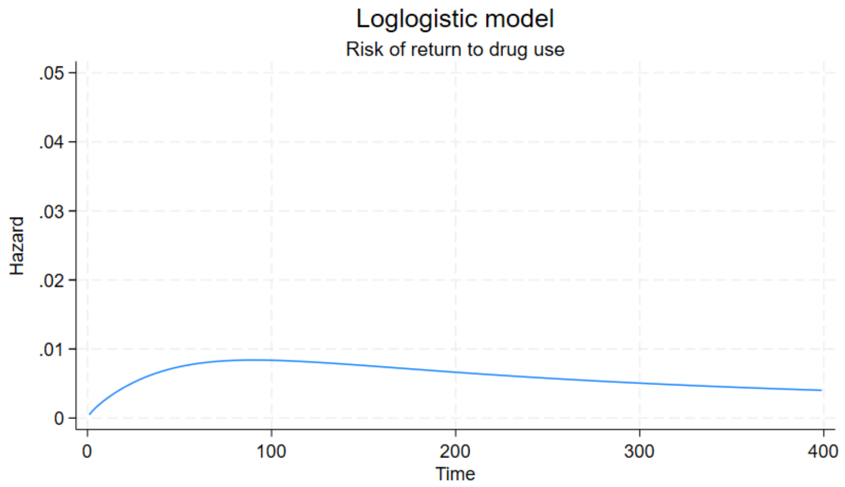


Parametric models I – model choice – Distribution of T

Log-logistic model

Model	Assumed shape of T
Cox	
Exponential	
Piece-wise constant exponential	
Gompertz	
Log-logistic	Increase AND decrease

- Increasing, then decreasing - Non-monotonic
- Shape parameter determine when the curve bends, a scale parameter determine the intensity

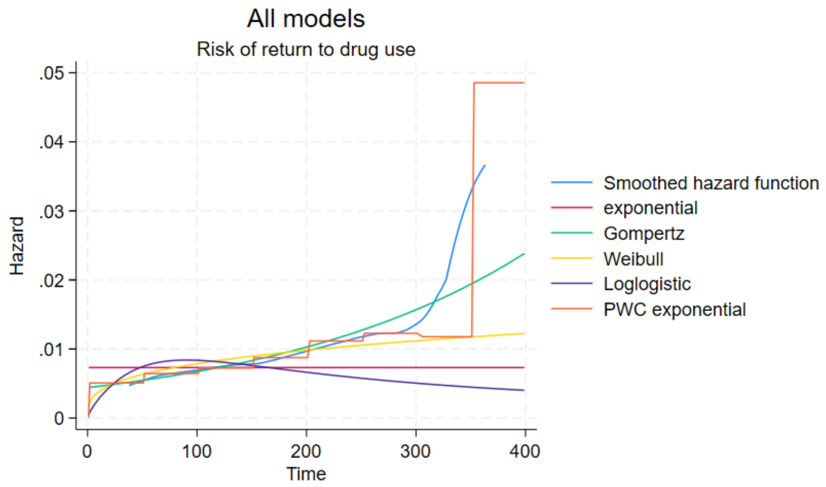


Parametric models I – model choice – Distribution of T

Summary so far:

Model	Assumed shape of T
Cox	None
Exponential	Constant
Piece-wise constant exponential	Any shape across intervals
Gompertz	Increasingly increasing
Log-logistic	Increase AND decrease

- Models differ in their assumed shape in the distribution of T
- Chose a model that fits your time process by:
  - Knowledge of about the studied phenomena
  - Contrast with your raw data
  - Comparison of model fit (labs)



Key estimates in EHA

1. Model assumptions about the shape of the hazard ✓
2. The proportional hazards assumption
3. Interpretation of estimates

15 minute break

Parametric models I – Proportional hazard assumption

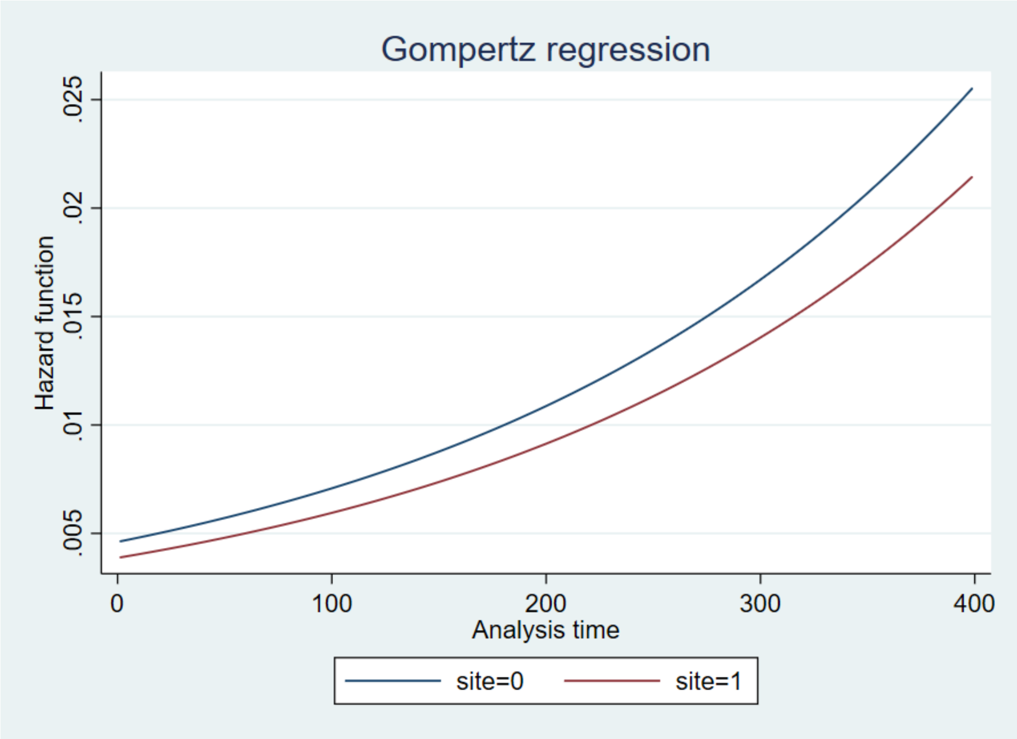
The proportional hazard assumption:

- One baseline hazard assumed for the full population
- The hazard rate for a covariate is scaled to that baseline.
- The proportion of hazard rates between values of covariates assumed identical across time.
- If violated, can bias estimates.
- Key for interpreting estimates of covariates.

Parametric models I – Proportional hazard assumption

The proportional hazard assumption (Example: hazard of return to drug use)

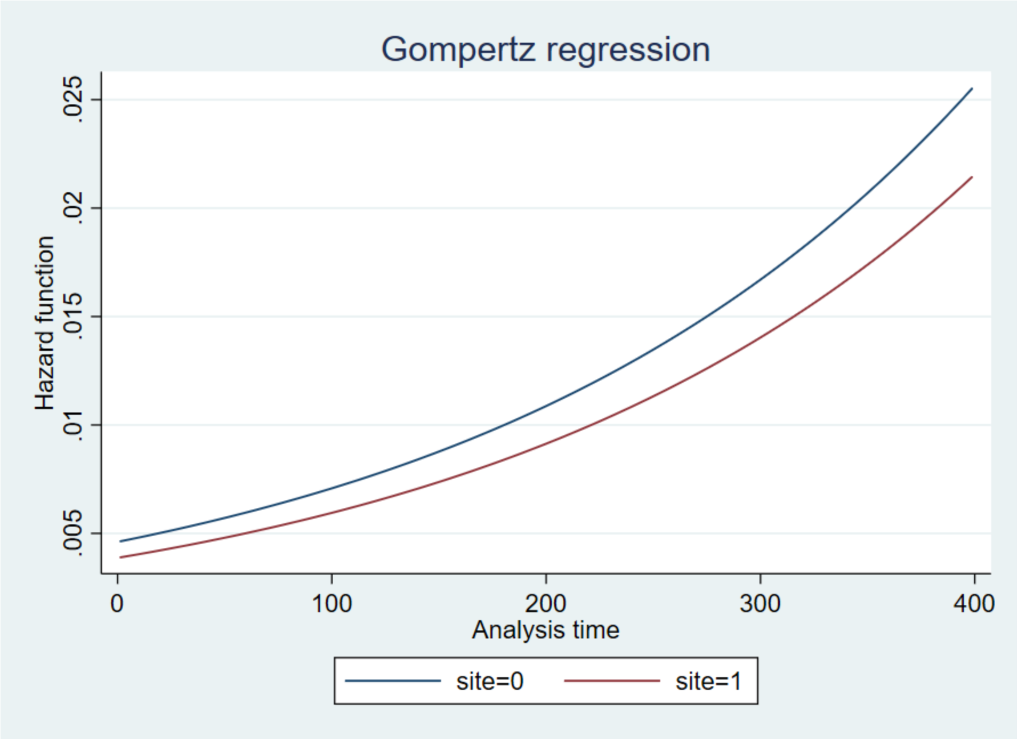
- Here, the hazard function of site of treatment
- h(t) differ by site of treatment...
- but proportionally so across the time period.
- Here, women always twice that of men



Parametric models I – Proportional hazard assumption

The proportional hazard assumption (diagnosis)

- Test statistics (Schoenfeld residuals test)
- Contrast to non-parametric model



Parametric models I – model choice – Distribution of T

Model	Prop. Hazard assumtion
Cox	Yes
Exponential	Yes
Piece-wise constant exponential	Yes within interval
Gompertz	Yes
Log-normal, Log-logistic	Not applicable

Key estimates in EHA

1. Model assumptions about the shape of the hazard ✓
2. The proportional hazards assumption ✓
3. Interpretation of estimates

Interpretation of estimates

- Log-hazard, hazard ratio
  - Most common form out estimates for continuous time models
  - Log-hazard and hazard ratio equivalent
  - Pertains to the hazard.
  - Covariate effect: higher/lower risk of event

- Accelerated failure time (AFT)
  - Possible to estimate for Exp and Weibull models continuous time models
  - Log-survival and Time ratio equivalent
  - Pertains to survival.
  - Covariate increase: longer/shorter duration until event

- Note the difference in direction of interpretation
  - longer duration until event vs. higher risk of event

Interpretation of estimates – log-hazard & hazard ratio

	Log hazard	Hazard ratio
rationale	B of exponential model	Most commonly used.
formal definition	the log of the hazard	exponent of the log of the hazard
formula	ln $h(t) = \beta 0 + \beta 1x1$	$h(t) = h$ * exp($\beta 1x1$)
range	– ∞ to + ∞	0 to +∞
Effect direction of B	0 = none < 0 = lower hazard > 0 = higher hazard	1 = none < 1 = lower hazard > = 1 higher hazard
interpretation	B of .99 indicates a increase of the log-hazard by .99 for ever unit increase of X	HR of 1.05 indicate a 5% higher risk relative to the reference category (categorical) HR of 1.05 indicate a 5% higher risk for every unit increase of X (continous)

Interpretation of estimates – log-hazard & hazard ratio

- Hazard ratios with and without model covariates
- Mode1 (Empty model):
  - Constant = underlying baseline hazard of the entire population.
- Model 2 (adjustment for sex) male = 1 (reference), female = 0:
  - Constant = hazard for reference group (men)
  - Female = hazard ratio for women relative men: the ratio of female to male hazards.

	Model 1 (Empty model)	Model 2 (adjust for sex)	Estimate for female hazard
Covariates		*Female-to-male hazard ratio*	
Female	-	1.5	1.5*.075 = .125
Constant	0.5	.075	
	Baseline hazard	*Male hazard*	*Female hazard*

Interpretation of estimates – log-hazard & hazard ratio

	AFT - Log survival	AFT Time ratio
rationale	B of exponential model	Most commonly used.
formal definition	the log of the survival time	exponent of the log of the survival time
range	– ∞ to + ∞	0 to +∞
Effect direction of B	0 = none < = 0 shorter duration > = 0 longer duration	1 = none < 1 = shorter duration. > 1 = longer duration
interpretation	B of .99 indicates a increase of the log survival duration by .99 for ever unit increase of X	TR of 1.05 indicate a 5% longer duration of survival compared to the reference category (categorical) TR of 1.05 indicate a 5% longer duration of survival for each unit increase of X (continous)

Interpretation of estimates – examples

ARTICLE

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-020-18926-3

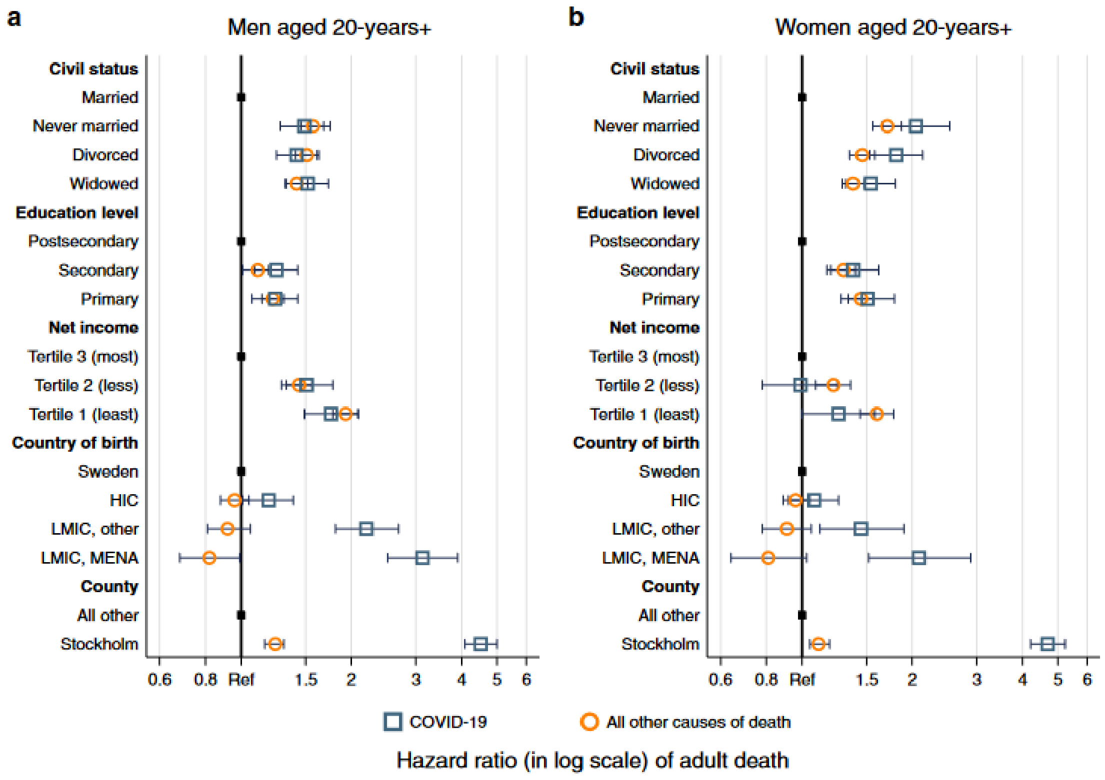


Fig. 1 Hazard ratios of dying from COVID-19 and all other causes of death for men and women in Sweden. a Men aged 20 years and older, error bars representing 95% confidence intervals of hazard ratios, $n = 3,876,881$ men. **b** Women aged 20 years and older, error bars representing 95% confidence intervals of hazard ratios, $n = 3,898,173$ women. Blue squares indicating COVID-19 mortality. Orange circles indicating mortality from all other causes of death.

Source: Drefahl, S., Wallace, M., Mussino, E., Aradhya, S., Kolk, M., Brandén, M., Malmberg, B., & Andersson, G. (2020). A population-based cohort study of socio-demographic risk factors for COVID-19 deaths in Sweden. Nature Communications, 11(1), 5097. https://doi.org/10.1038/s41467-020-18926-3

Interpretation of estimates – examples

Table 11: Hazard Model Estimates for Fertility Decisions

	Couples with 1+ Child		Couples with 2+ Children	
Dep. Variable	Older Cohorts	Young Cohorts	Older Cohorts	Young Cohorts
First born is Male	0.8984*	0.8328***		
	(0.0510)	(0.0417)		
Two Boys			0.4985***	0.4783***
			(0.0478)	(0.0580)
Boy and Girl			0.5438***	0.5261***
			(0.0406)	(0.0534)
Spouse Birth Control Pref. (ref. Encouraged)				
Opposed	1.0514	1.0098	1.4985 ***	1.0406
	(0.1600)	(0.1106)	(0.2124)	(0.1903)
Showed no enthusiasm	0.8834	1.0062	1.0772	1.0940
	(0.7787)	(0.0748)	(0.1030)	(0.1435)
Test (χ^2)			0.90	0.71
$Prob > \chi^2$			0.3427	0.4005
Observations	2,032	3,062	1,855	1,778

Notes: Robust standard errors are in parenthese. *** p<0.01, ** p<0.05, * p<0.1. The table reports the *hazard ratios* for the birth of a second child (for couples with at least one child) and a third child (for couples with at least two children). A *Weibull* distribution is assumed for the baseline hazard. "Older Cohort" refers to the sample of respondents aged 32-39. "Young Cohort" refers to the sample of respondents aged 20-31. Each equation includes controls (not shown) for the mother's current age, the mother's age at marriage, the age gap within the couple, number of years of marriage, the couples' years of schooling, and the grandparents' socio-economic status, and district fixed-effects. "Test (χ^2)" refers to the chi-squared test statistic for a test of the equality of the coefficients of "Two Boys " and "Boy and Girl". The p-values are shown below the test statistics.

Source: Asadullah, M. N., Mansoor, N., Randazzo, T., & Wahhaj, Z. (2021). Is son preference disappearing from Bangladesh? *World Development*, *140*, 105353. https://doi.org/10.1016/j.worlddev.2020.105353