Survival Analysis and Event History

Andy Grogan-Kaylor

23 Jul 2020

# Introduction

“Survival analysis is a key technique in data-driven decision-making, which is now central to public interest because of COVID-19. Applying the correct technique for the specific question at hand is crucial for credible public health inferences. If you are interested in assessing how a risk factor or a potential treatment affects the progression of a disease—such as how long a patient takes to recover—then survival analysis techniques come into play. Survival analysis deeply respects the ultimate source of its data, often the disease experience or even the life and death of human patients. It seeks to exploit every last drop of information that this experience can render for saving lives—in particular, not only whether patients survived, but how long, and why. And it strives to do so with minimal assumptions, so that the data are truly driving the decision.”

—SAS Corporation

# Key Concepts

WHO CARES how we measure time? Isn’t it self-evident?

* Implementations differ; formulas are our friends
* : formula (effect on hazard)

# The Hospital Bed Problem

* Imagine a *Hypothetical Hospital* 🏥
* Imagine that there are 52 patients *total*.

⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤

* 51 of the patients are *long term patients*, who each stay for *1 year*.

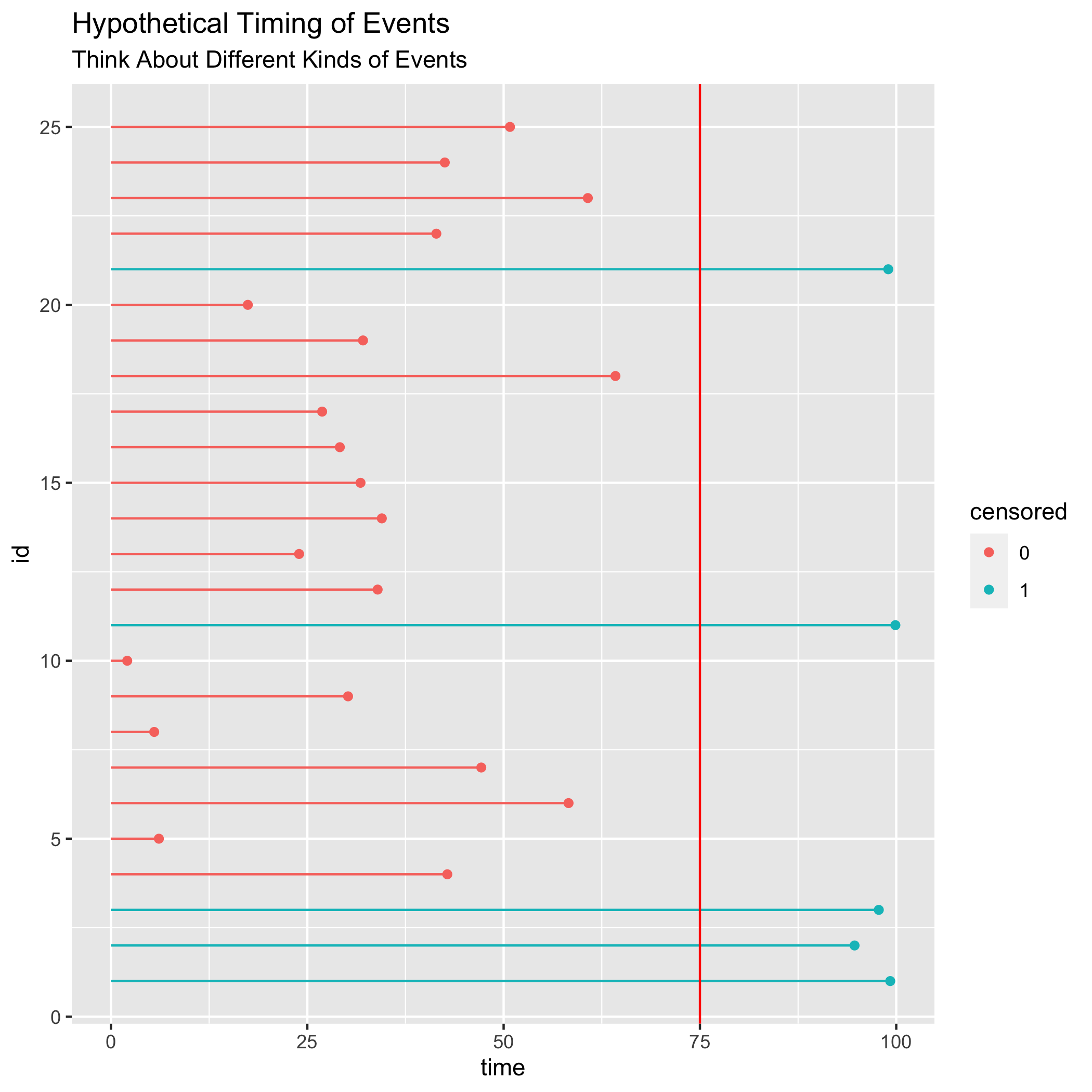
⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤

* 1 of the patients is a *short term patient*, who stays for *1 week*.

⬤

Is this a hospital that serves mostly long-term, or short term patients?

# How To Measure Length of Stay (1)



## Animated

See [times-events-and-censoring.html](./times-events-and-censoring.html)

# How To Measure Length of Stay (2)

* Event happened within a specified time (yes/no)
  + Statistically accurate, but we lose information on *when* the event happened.
  + Statistically *less efficient*.
* Time until Event
  + What to do with events that haven’t happened yet? (Censoring)
  + Code as NA. Loss of information. Possible bias.
  + Code as 0. Possible bias. They might happen at some point.
  + Code as time of censoring. Possible bias. They might never happen. They might happen much later.

# A Policy Example (Welfare Reform, 1996)

From LaDonna Pavetti (1995)

* time in months
* new entrants (percent)
* all current recipients at a point in time (percent)

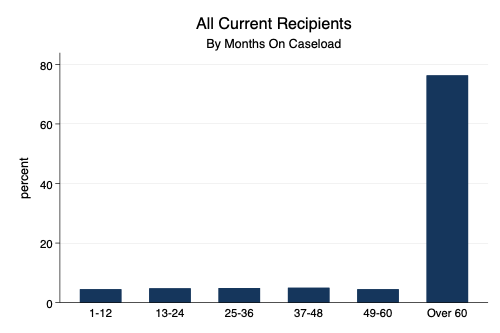
. clear all

. use Pavetti.dta  
(Written by R. )

. list, abbreviate(25) // list out the data  
  
 ┌─────────────────────────────────────────────────┐  
 │ time new\_entrants all\_current\_recipients │  
 ├─────────────────────────────────────────────────┤  
 1. │ 1-12 27.4 4.5 │  
 2. │ 13-24 14.8 4.8 │  
 3. │ 25-36 10 4.9 │  
 4. │ 37-48 7.7 5 │  
 5. │ 49-60 5.5 4.5 │  
 ├─────────────────────────────────────────────────┤  
 6. │ Over 60 34.6 76.3 │  
 └─────────────────────────────────────────────────┘

. graph bar (asis) all\_current\_recipients, /// this particular set of options was difficu  
> lt to figure out!  
> over(time) ///  
> title("All Current Recipients") ///  
> sub("By Months On Caseload") ///  
> ytitle("percent") ///  
> scheme(michigan)

. graph export all\_current\_recipients.png, width(500) replace  
(file all\_current\_recipients.png written in PNG format)

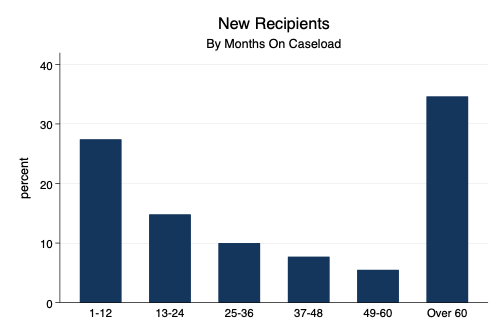


All Current Recipients by Months on Caseload

# Welfare Reform (2)

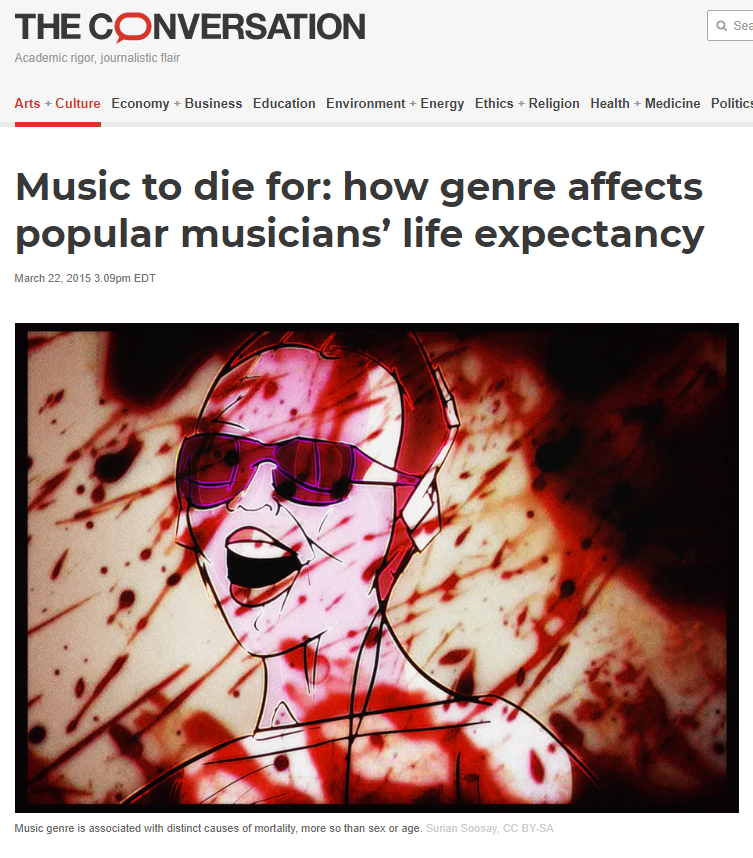
. graph bar (asis) new\_entrants, ///  
> over(time) ///  
> title("New Recipients") ///  
> sub("By Months On Caseload") ///  
> ytitle("percent") ///  
> scheme(michigan)

. graph export new\_recipients.png, width(500) replace  
(file new\_recipients.png written in PNG format)



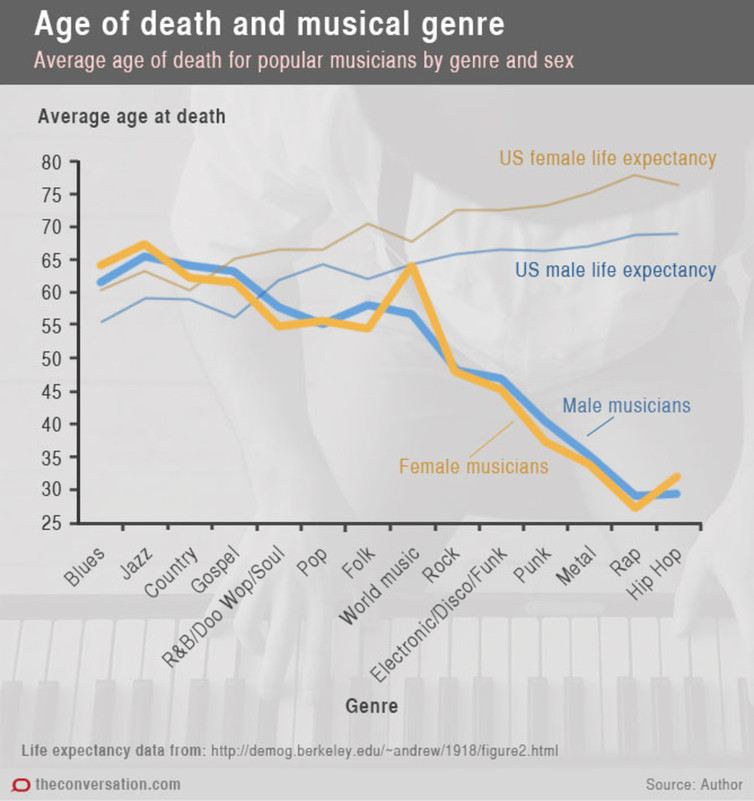
New Recipients by Months on Caseload

# Musicians and Mortality (1)



Music To Die For

# Musicians and Mortality (2)



Musician Mortality

# Cox Proportional Hazards Model

# Formula

the rate of occurrence.

We don’t directly estimate the hazard, but estimate the effect of covariates on the hazard.

The event (birth, death, program entry, program departure) is coded as 1, so we are estimating the association of the covariates with event occurrence.

# Cox Proportional Hazards Model in Stata

. clear all

. webuse drugtr // demonstration data set from Stata  
(Patient Survival in Drug Trial)

## Setup of Data

. stset // show st setup of data  
-> stset studytime, failure(died)  
  
 failure event: died != 0 & died < .  
obs. time interval: (0, studytime]  
 exit on or before: failure  
  
──────────────────────────────────────────────────────────────────────────────  
 48 total observations  
 0 exclusions  
──────────────────────────────────────────────────────────────────────────────  
 48 observations remaining, representing  
 31 failures in single-record/single-failure data  
 744 total analysis time at risk and under observation  
 at risk from t = 0  
 earliest observed entry t = 0  
 last observed exit t = 39

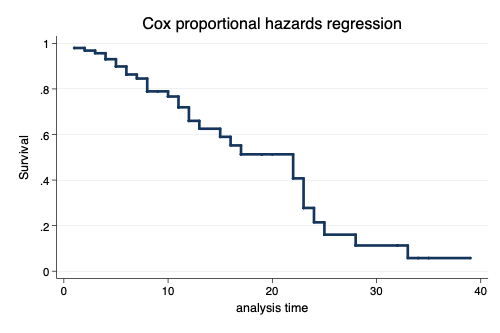
## Cox Proportional Hazards Model

. stcox age drug // run Cox Proportional Hazards Model  
  
 failure \_d: died  
 analysis time \_t: studytime  
  
Iteration 0: log likelihood = -99.911448  
Iteration 1: log likelihood = -83.551879  
Iteration 2: log likelihood = -83.324009  
Iteration 3: log likelihood = -83.323546  
Refining estimates:  
Iteration 0: log likelihood = -83.323546  
  
Cox regression -- Breslow method for ties  
  
No. of subjects = 48 Number of obs = 48  
No. of failures = 31  
Time at risk = 744  
 LR chi2(2) = 33.18  
Log likelihood = -83.323546 Prob > chi2 = 0.0000  
  
─────────────┬────────────────────────────────────────────────────────────────  
 \_t │ Haz. Ratio Std. Err. z P>|z| [95% Conf. Interval]  
─────────────┼────────────────────────────────────────────────────────────────  
 age │ 1.120325 .0417711 3.05 0.002 1.041375 1.20526  
 drug │ .1048772 .0477017 -4.96 0.000 .0430057 .2557622  
─────────────┴────────────────────────────────────────────────────────────────

## Graph Survival Curves

. stcurve, survival scheme(michigan) // survival curve

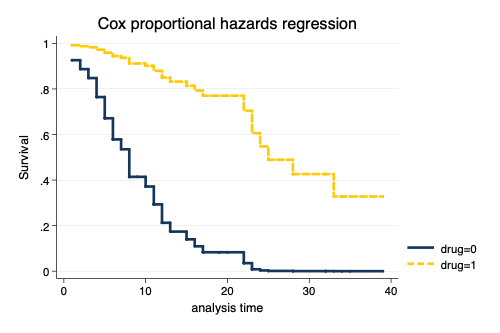
. graph export survival1.png, width(500) replace  
(file survival1.png written in PNG format)



Survival Curve

. stcurve, survival at1(drug=0) at2(drug=1) scheme(michigan) // survival curve by group

. graph export survival2.png, width(500) replace  
(file survival2.png written in PNG format)



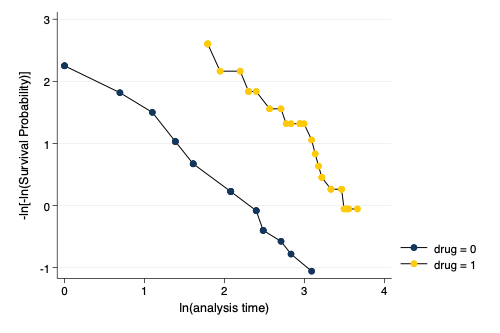
Survival Curve by Drug Group

# Proportional Hazards Assumption

. estat phtest // formal test of PH assumption  
  
 Test of proportional-hazards assumption  
  
 Time: Time  
 ────────────┬───────────────────────────────────────────────────  
 │ chi2 df Prob>chi2  
 ────────────┼───────────────────────────────────────────────────  
 global test │ 0.43 2 0.8064  
 ────────────┴───────────────────────────────────────────────────

. stphplot, by(drug) scheme(michigan) // graphical test of PH assumption  
  
 failure \_d: died  
 analysis time \_t: studytime

. graph export ph.png, width(500) replace  
(file ph.png written in PNG format)



Graphical Assessment of Proportional Hazards Assumptions

# Unobserved Heterogeneity