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*.

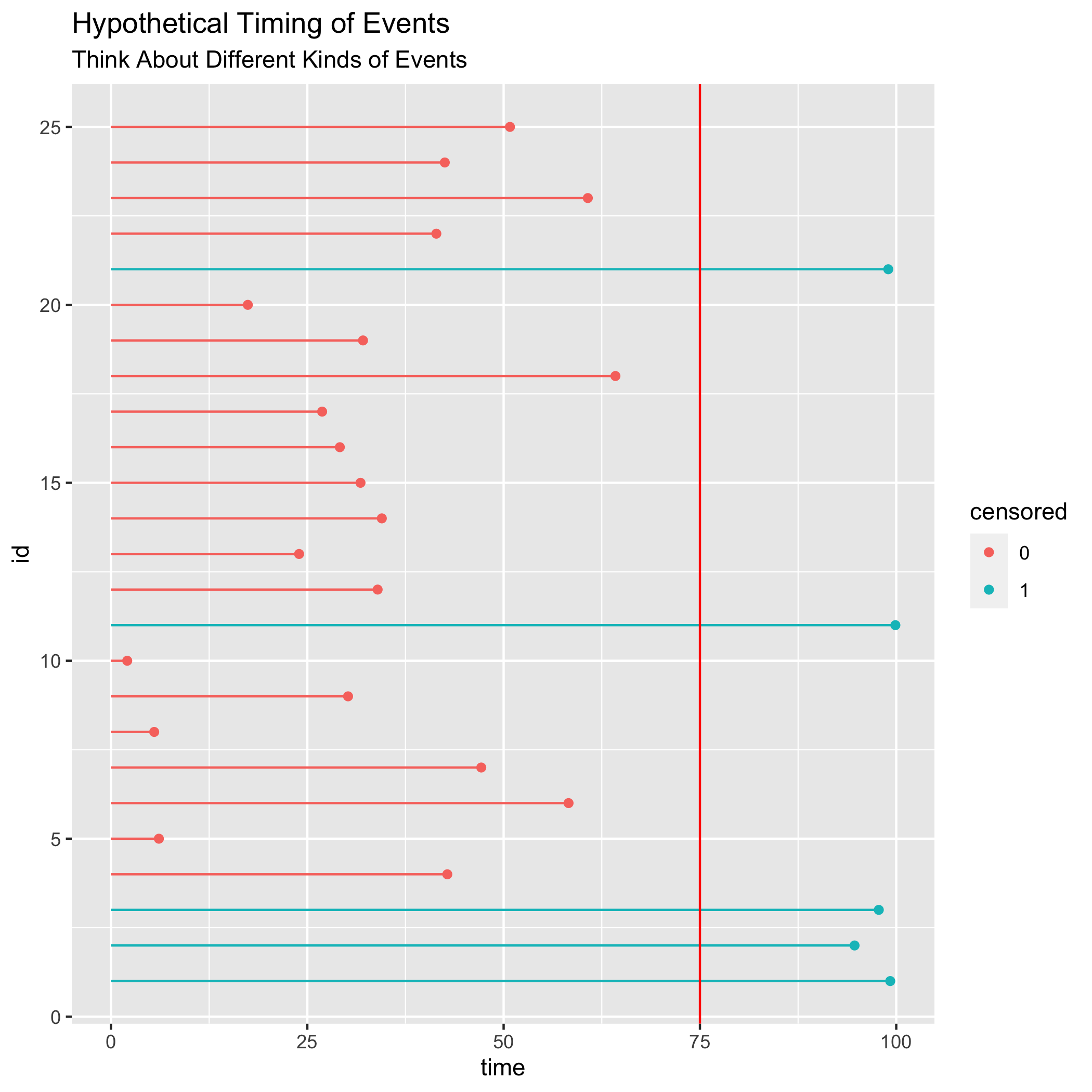
⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤ ⬤

* 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 │  
 └─────────────────────────────────────────────────┘