



● Your **First Look** Into Survival Analysis







Presented by:

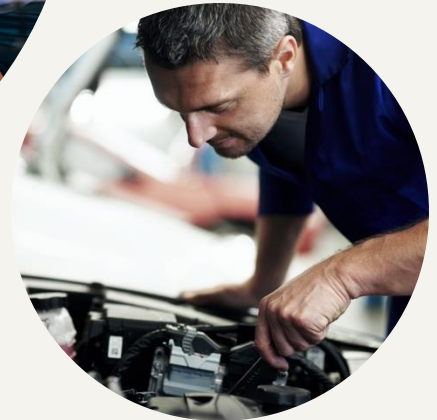
Emmanuel DJEGOU

PhD Candidate in Statistics @Missouri S&T

Data Intern @RGA

What is Survival Analysis?

- **Study of time until an event occurs**
- Events include:
 -  **Death**
 -  **System failure**
 -  **Relapse**
 -  **Customer churn**
 -  **Morbidity**
 -  **Retirement**
- Crucial in **medicine, engineering, business, insurance**, and beyond



Key Variables

Event Time (T):

Time until the event occurs.

Censoring (C):

Occurs when the event is not observed by the end of the observation period.

Status Indicator (δ):

- $\delta=1$: Event observed.
- $\delta=0$: Event censored.
- **Data: ($Y = \min(T, C)$, δ):** Observed time is $Y=\min(T,C)$, with indicator δ .

Study Start



Event



Study End



Timeline

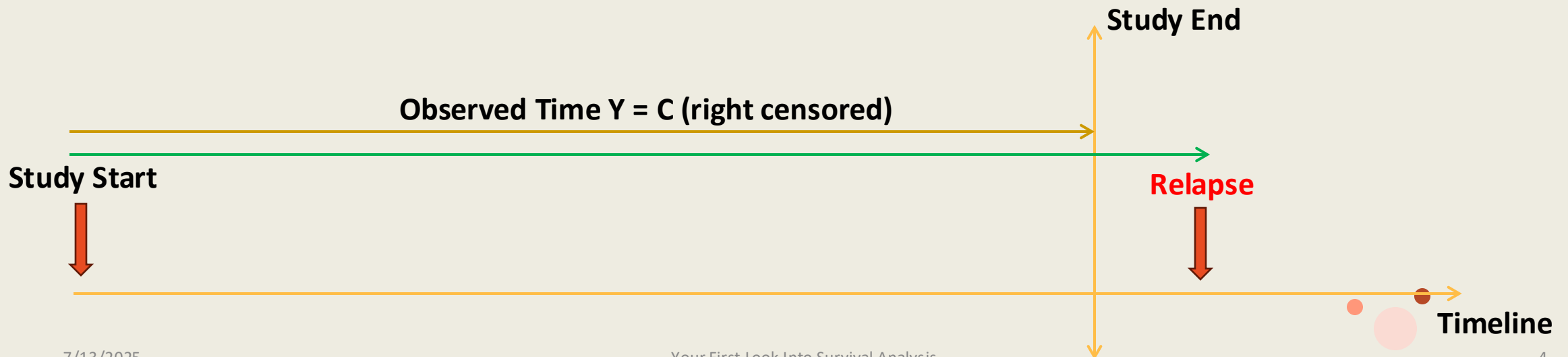


Types of Censoring

- **Right censoring** (most common): event has not occurred *yet*

Example:

- Patient A enrolls in the study and is followed for the full 5 years without **relapsing**.
- Their relapse time is right-censored at 5 years.
- Why? Because we only know that their relapse time is at least 5 years, but we don't know the exact time of relapse (if any).

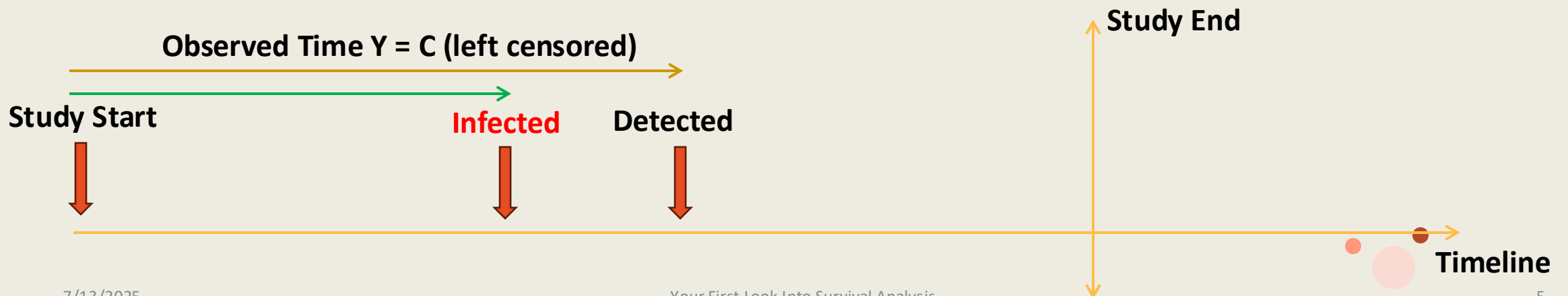


Types of Censoring

- **Left censoring**: event occurred before observation

Example:

- A patient comes in for their first test, and the results show they are already **infected**.
- However, there is no record of when the infection actually occurred.
- The infection happened before the first observation, but the exact time of infection is unknown—only that it occurred before detection.

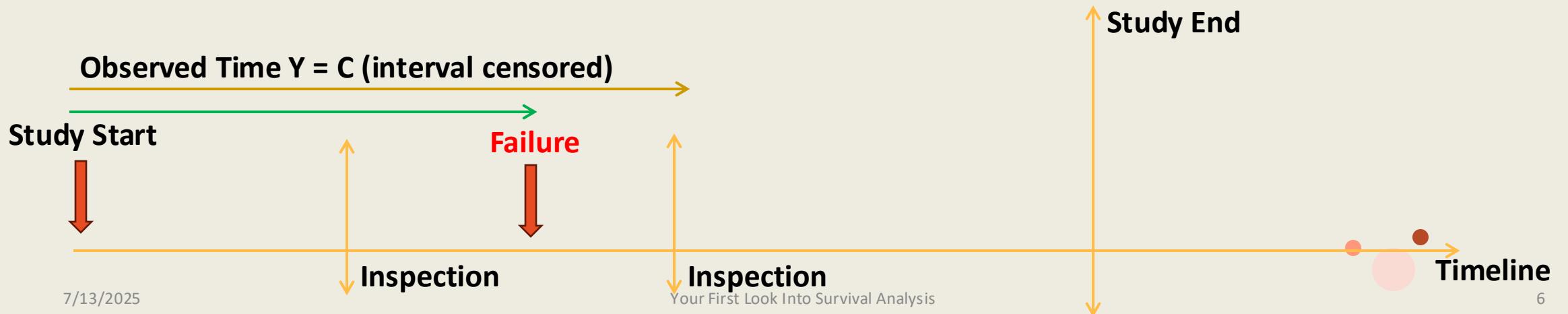


Types of Censoring

- **Interval censoring:** event occurred in a time interval

Example:

- A company monitors machines for component failure with routine inspections every 100 hours.
- A machine's component is functioning at 500 hours,
- But at the next inspection at 600 hours, the component is found to have already failed.
- The failure occurred between 500 and 600 hours, but the exact time is unknown.
- That's interval censoring — the failure time is known only to lie within a time interval.



Survival & Hazard Functions

- Definition:

$$S(t) = P(T > t) = 1 - F(t)$$

- $F(t)$ is the CDF.
- The probability an individual survives past time t .
- Starts at 1, decreases over time
- If $f(t)$ is the PDF:

$$S(t) = \int_t^{\infty} f(u) \, du = 1 - F(t)$$

- Or:

$$f(t) = -\frac{dS(t)}{dt}$$

Survival & Hazard Functions

- Hazard Function:

$$\lambda(t) = \frac{f(t)}{S(t)}$$

- Think of it as: *"If you're alive at time t , what's the risk you die instantly?"*
- *Cumulative Hazard function:*

$$\Lambda(t) = \int_0^t \lambda(u) du$$

- Relation to survival:

$$S(t) = \exp(-\Lambda(t))$$

- It is the total accumulated risk of experiencing the event up to time t .
- *Think it as the "risk exposure" that builds up over time.*
- The longer you "survive," the more risk you've accumulated — but not necessarily experienced yet.

Want to Learn More?



For an in-depth treatment, see:

- ✓ **Collett, D. (2023).** *Modelling Survival Data in Medical Research* (4th ed.). CRC Press, Taylor & Francis Group.

📌 Follow Me For More Exciting Discussions On Data Science, Statistics And Survival Analysis

[All My Channels In One Place](#)

Feel Free To Connect, Collaborate, Or Just Say Hi! 🤝

Thank You!