HOLMUSK

Data Challenge

Javier Manzano

08/04/2021

Agenda

1. Problem
2. Approach to problem solving
3. Results
4. Insights

Agenda

1. **Problem**
2. Approach to problem solving
3. Results
4. Insights

1. Situation (i)

A

B

Bleeding Event

1. Information (ii)

Patients

- Socio-demographics
- Diagnosis
- Lab Values
- Therapies
- …

A

B

Drugs

- Treatment
- Time
- Bleeding Event

1. Objective & Problem (iii)

- Objective:
  - To compare the **efficacy of the two drugs** (A and B) in terms of the risk of adverse events

- Problem:
  - It is **not a clinical trial**
  - The two groups **may not have balanced patient characteristics**
    - It could **mislead results/insights/decisions**

- How to reduce this problem?
  - It is necessary to identify **relevant variables** and **balance the patient characteristics**
    - Cox Model and Propensity Score Matching
    - Survival Analysis

Agenda

1. Problem

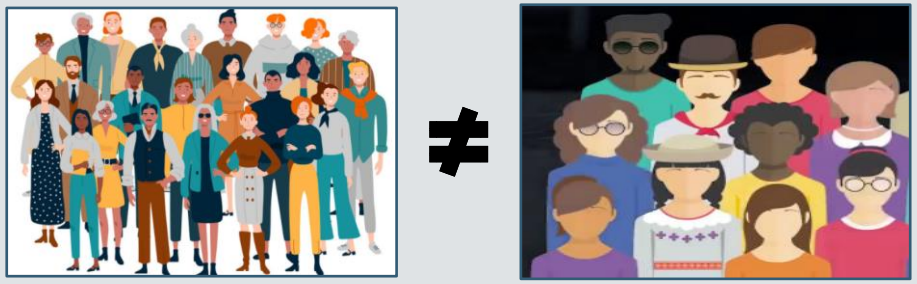
2. Approach to problem solving

3. Results

4. Insights

2. Approach

- 1st Step: Unbalanced patient characteristics
  - Relevant variables by using Cox Model
  - Survival time (median) by using KM Method
  - Impacts from unbalanced data



- 2nd Step: Balanced patient characteristics
  - Relevant variables by using Cox Model
  - Randomized controlled trial by using PSM
  - Survival time (median) by using KM Method
  - Impacts from balanced data

Cox Model (CM), Kaplan-Meier Method (KM), and Propensity Score Matching (PSM)

Agenda

1. Problem

2. Approach to problem solving

3. **Results**

4. Insights

3. Results (i)

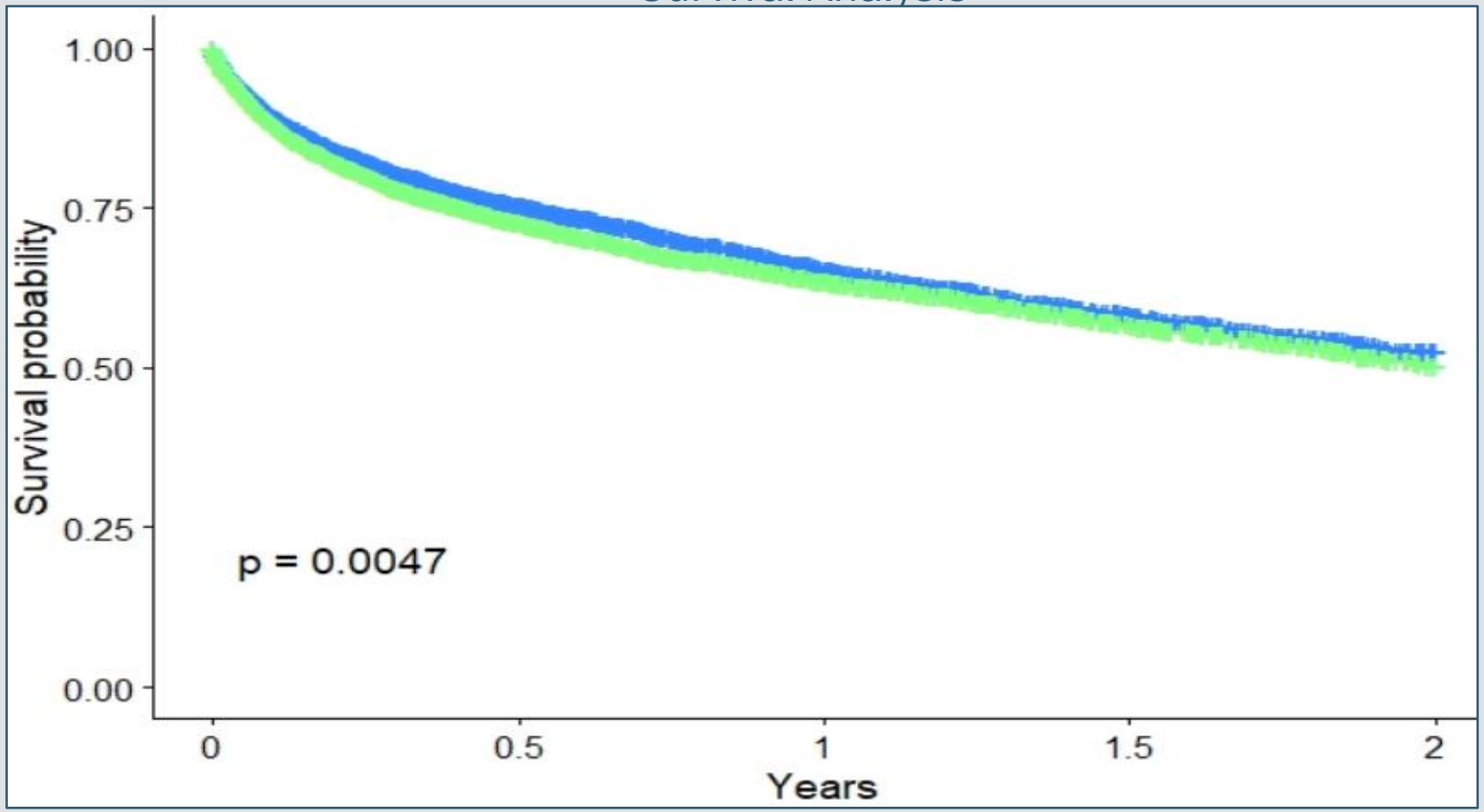
- Unbalanced patient characteristics

(17) Relevant variables*

Variables	Hazard Ratios
"age"	1.00
"other_drugs_2"	1.26
"other_drugs_3"	1.22
"other_drugs_8"	1.17
"diagnosis_1"	0.89
"diagnosis_4"	0.62
"diagnosis_6"	1.23
"diagnosis_8"	0.92
"diagnosis_9"	0.91
"diagnosis_10"	1.37
"diagnosis_11"	1.18
"diagnosis_12"	3.35
"diagnosis_13"	0.83
"diagnosis_14"	1.42
"lab_1"	1.09
"Diag_Score_1"	1.09
"Diag_Score_2"	1.08

* At 5%

Survival Analysis



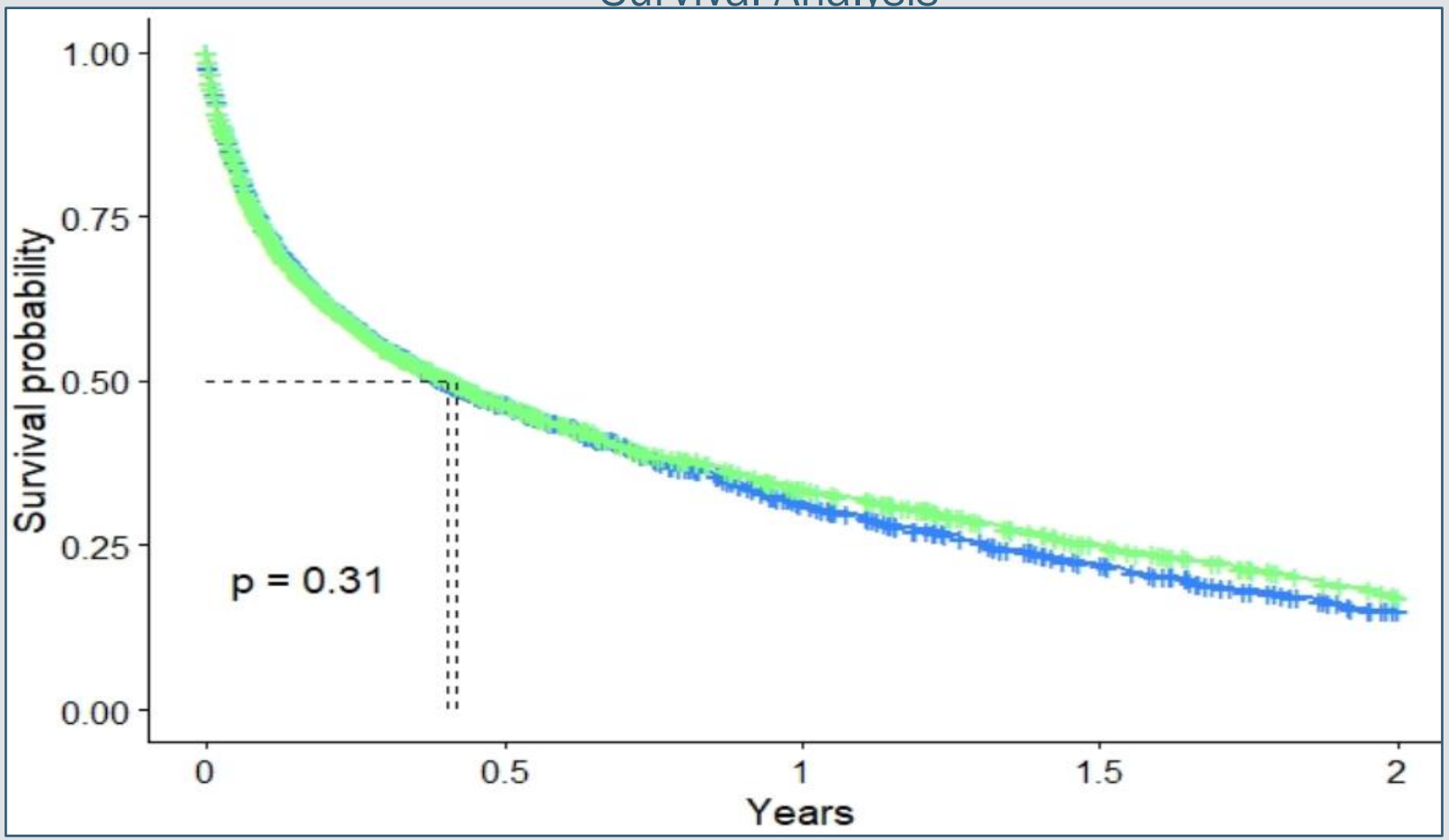
3. Results (ii)

- Balanced patient characteristics

(11) Relevant variables

	Variables	Hazard Ratios
At 5%	"other_drugs_2"	1.10
	"other_drugs_8"	1.26
	"diagnosis_9"	0.91
	"diagnosis_12"	1.40
	"diagnosis_14"	0.52
	"Diag_Score_1"	1.08
	"Diag_Score_2"	1.05
At 10%	"diagnosis_1"	0.93
	"diagnosis_4"	0.72
	"diagnosis_11"	1.16
	"lab_1"	1.03

Survival Analysis



Agenda

1. Problem

2. Approach to problem solving

3. Results

4. Insights

4. Insights (i)

- Unbalanced patient characteristics (UD):
  - There are 17 relevant variables
  - There is statistically significant difference between drugs
  - Questionable results

- Balanced patient characteristics (BD):
  - There are 11 relevant variables
  - There is no statistically significant difference
  - More accurate results

4. Insights (ii)

- UD vs BD
  - More variables = Complex model
  - Coefficients tend to be higher
  - It could mislead results/decisions

  - Less variables = Simple model
  - Coefficients tend to be moderate
  - It can leads better results/decisions

- Suggestion
  - To work with BD = Better results/decisions

Variables	Hazard Ratios	
	UD	BD
"diagnosis_4"	0.62	0.72
"other_drugs_8"	1.17	1.26
"diagnosis_1"	0.89	0.93
"diagnosis_9"	0.91	0.91
"diagnosis_12"	3.35	1.40
"diagnosis_14"	1.42	0.52
"other_drugs_2"	1.26	1.10
"diagnosis_11"	1.18	1.16
"lab_1"	1.09	1.03
"Diag_Score_2"	1.08	1.05
"Diag_Score_1"	1.09	1.08

HOLMUSK

Data Challenge

Javier Manzano

08/04/2021