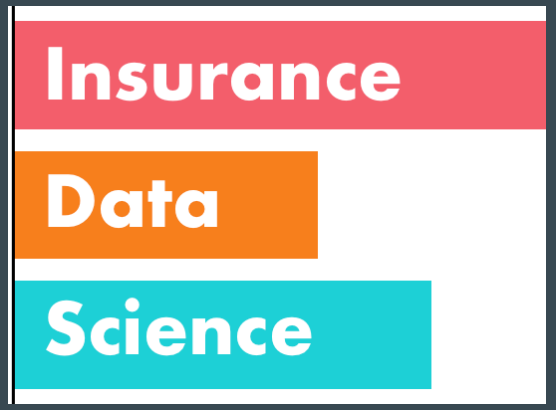
Process Mining Applied to Complex Medical Claims Management

Presented by Tim King

Insurance Data Science Conference
City, University of London
June 17, 2021



Business Problem

“By the time a fraudulent [healthcare] claim has been paid, the probability of recovering the money is approximately 1%”

Source: Boston Consulting Group

(Photo source: iStock)

“Approximately two-thirds of all claims we define as large losses started out as fairly routine claims, including back, shoulder and knee injuries.”

Source: Property Casualty 360

teradata.

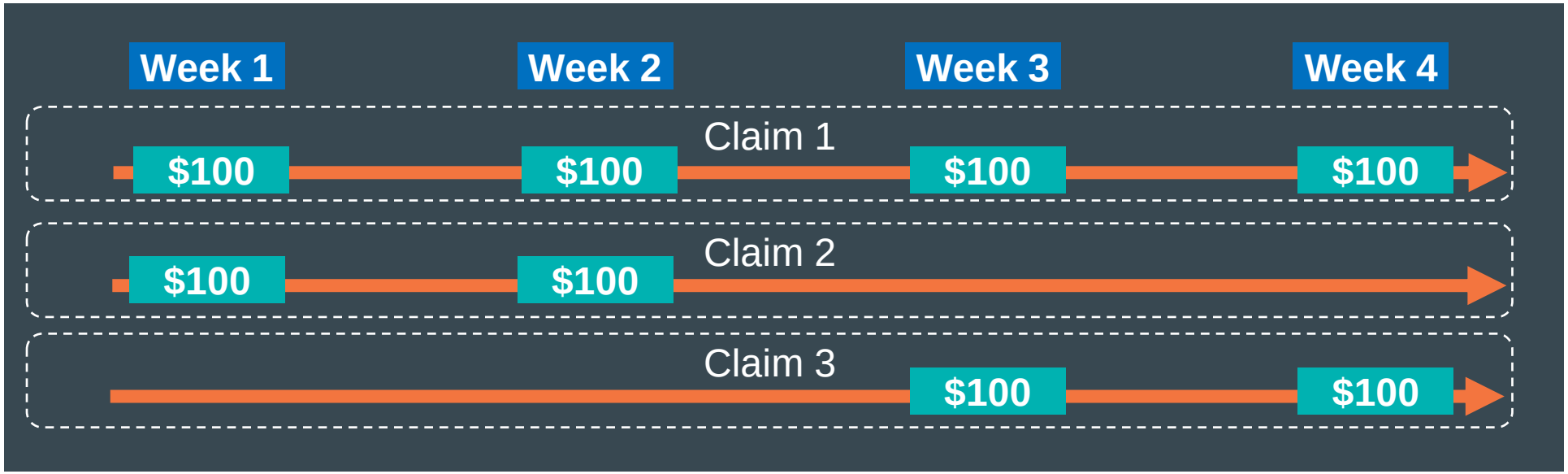
Why Process Mining

Claim Pathways are the New Frontier of Casualty Claims Analysis

LAGGARD
Bill-level validation

COMMON
Claim-level averages

LEADER
Claim pathways



teradata.

Two Key Questions

#1 Which claim lifecycles are anomalous?

#2 Which claim lifecycles lead to large $$ spend?

teradata.

The Approach – R Packages Designed Specifically for This Type of Problem

bupaR

"integrated suite of R-packages for the handling and analysis of business process data"

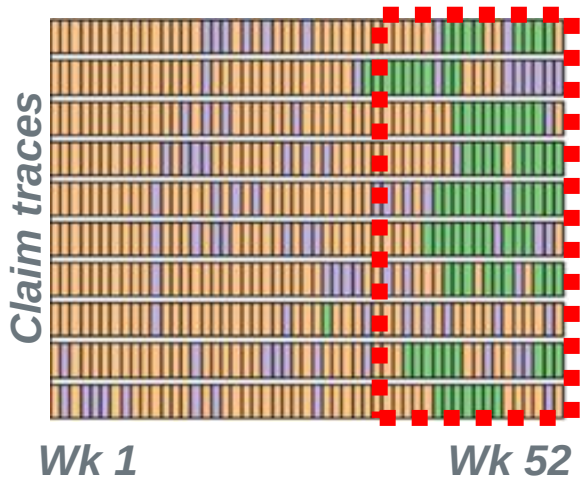
"Toolbox for the manipulation, description and rendering of sequences, and more generally the mining of sequence data"

What we Found – Anomaly Detection

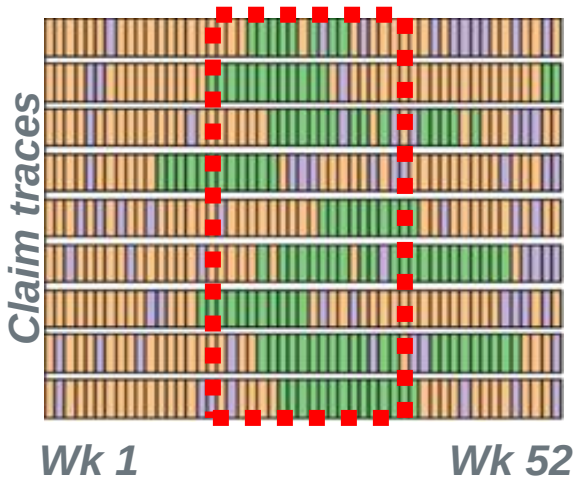
Unlisted Rehabilitation – *procedure*

CPT Code 97799 – "Unlisted Rehabilitation Service or Procedure"

Cluster 1 (n = 40)



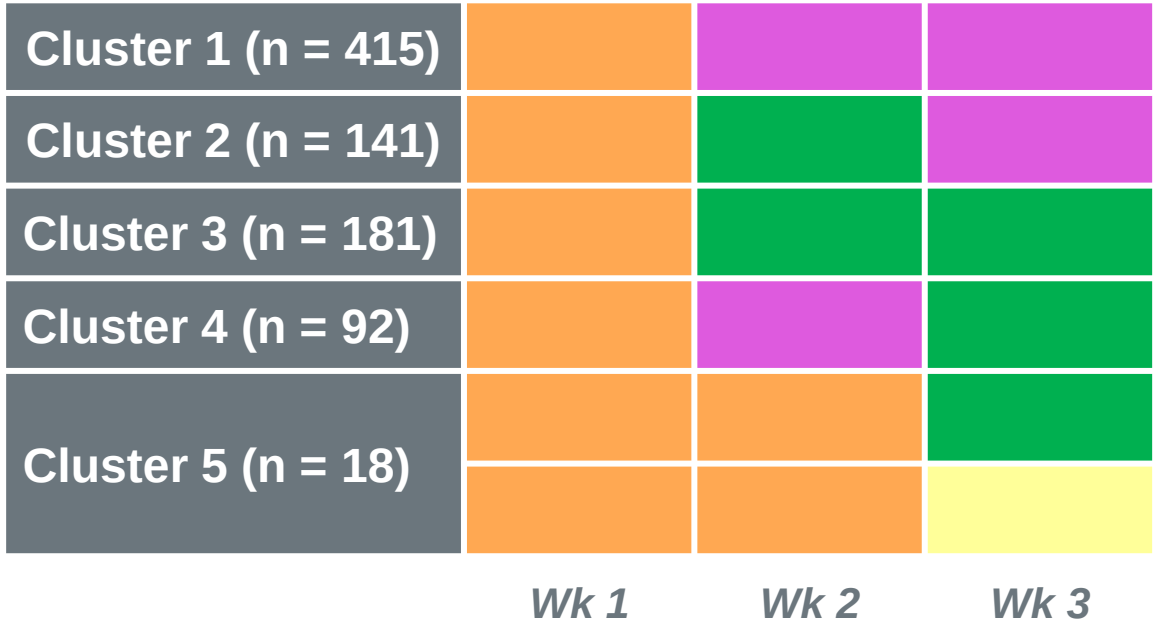
Cluster 2 (n = 9)



Other Procedure | No activity | 97799

"Other" Injury – *diagnosis*

ICD-10 Code T14.90X – "Injury, Unspecified"



T14.90X | No activity | Different Diagnosis

teradata.

What we Found – Large Loss Prediction

Diagnosis trajectory matters, illustrating using claims starting with S33.5XX

#1 Predictor of Large Loss Claims[1]:
Combined Nature and Location of Injury (e.g. "shoulder bursitis")

2. Age at injury
3. Incapacity description from the last known medical certificate
4. Specialty of most frequented provider
5. Combined nature and location of injury for the claim immediately prior to the current claim
6. …….
36. Proportion of predominant provider visits count to overall provider visits

Insurer	Cluster Relativities		
	(1) Same Diagnosis	(2) Same Chapter / No Activity	(3) Different Chapter
Travelers	1.00	1.54	2.70
Texas Mutual	1.00	1.61	1.84
Texas Municipal League	1.00	2.09	3.40

1. Diagnosis is constant for the first three weeks of the claim
2. In weeks two and three, there is either no activity or a different diagnosis that is in the same chapter
3. Neither week two nor week three has the initial chapter

[1]Source: "Analytics-assisted triage of workers' compensation claims". Ivan Lebedev, Inna Kolyshkina, Marcus Brownlow, Colin Khoo

teradata.

Conclusions and Next Steps

- Scale limitations
  - If overcome, can model on more data
  - Opportunities to do in-database, especially distance calculation
- Opportunities to improve the ‘math’
  - Alternative distance methods and / or cost functions
  - Improved intelligence for selecting cluster count
- Additional use cases – e.g., actuarial
  - Reserving and finance
  - Pricing & underwriting

teradata.

Contact the Presenter

Tim King
Insurance Industry Consultant
+1 301-520-9344
tim.king@teradata.com
LinkedIn

teradata.

Thank you.

teradata.

©2021 Teradata