Dynamic Hierarchical Aggregation, Selection Bias Tracking, and Detailed Subset Comparison for High-Dimensional Event Sequence Data

Jonathan Zhang, David Borland, Wenyuan Wang, Joshua Shrestha, David Gotz
University of North Carolina at Chapel Hill

Introduction

- *The collection and analysis of **event sequence data** has become more common over a wide range of domains*
- *We can use **visual analytics** to gain insights about this data*
- ***Medical data analysis** is a common application due to:*
  1. *The large amount of event sequence data*
  2. *The variety of population-based questions that this form of data can help answer*

The Problem

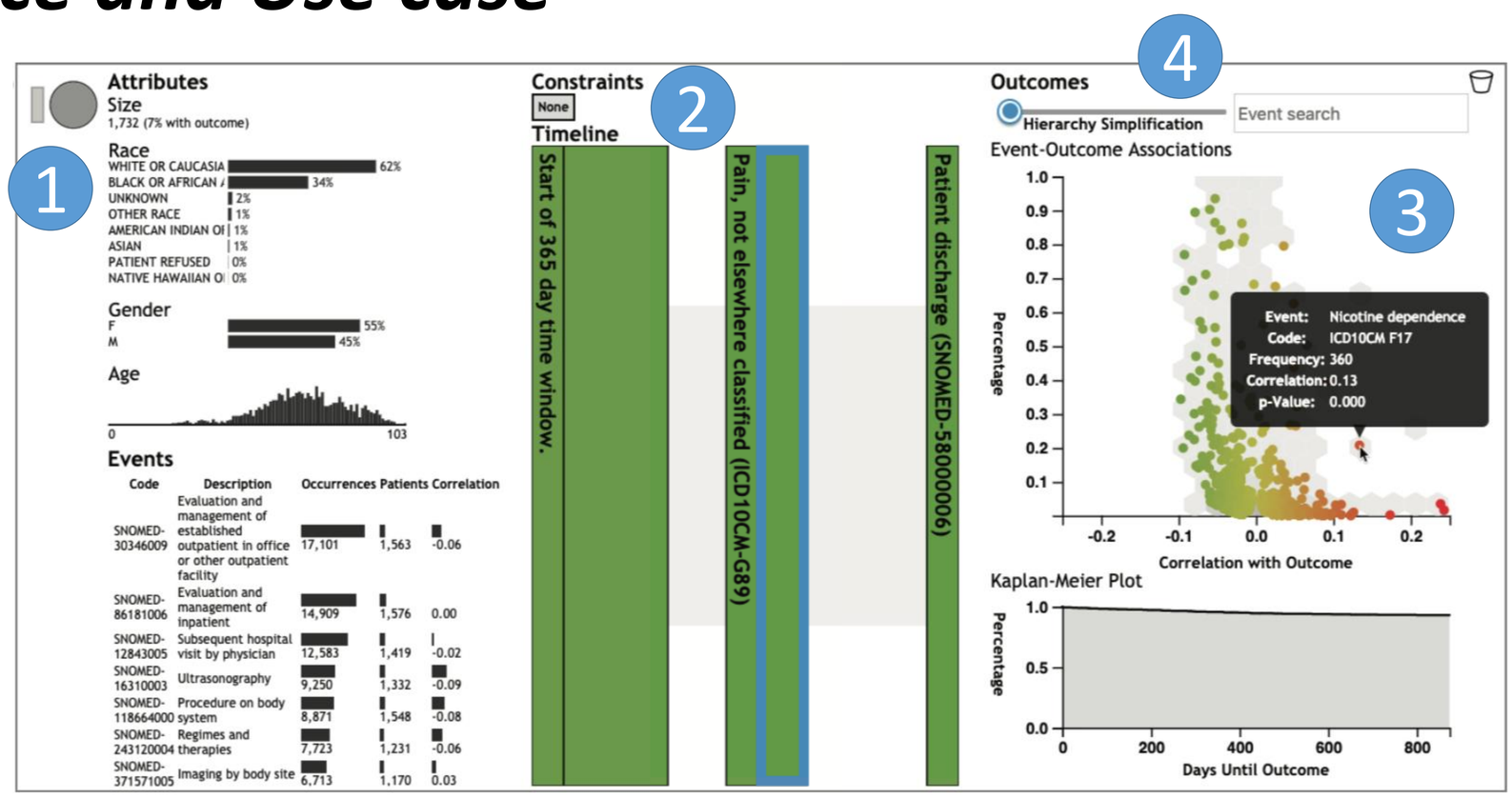
- *During data analysis, common problems arise including:*
  1. *The **large number of events types** prevents users from effectively interpreting the data*
  2. *The creation of subsets through interaction can introduce **selection bias***
- *Cadence is a visual analytics system designed to help overcome these challenges*

Visualization of Cadence and Use case

Timeline Panel

User selects patients who were discharged from the hospital after being previously diagnosed with any form of pain.

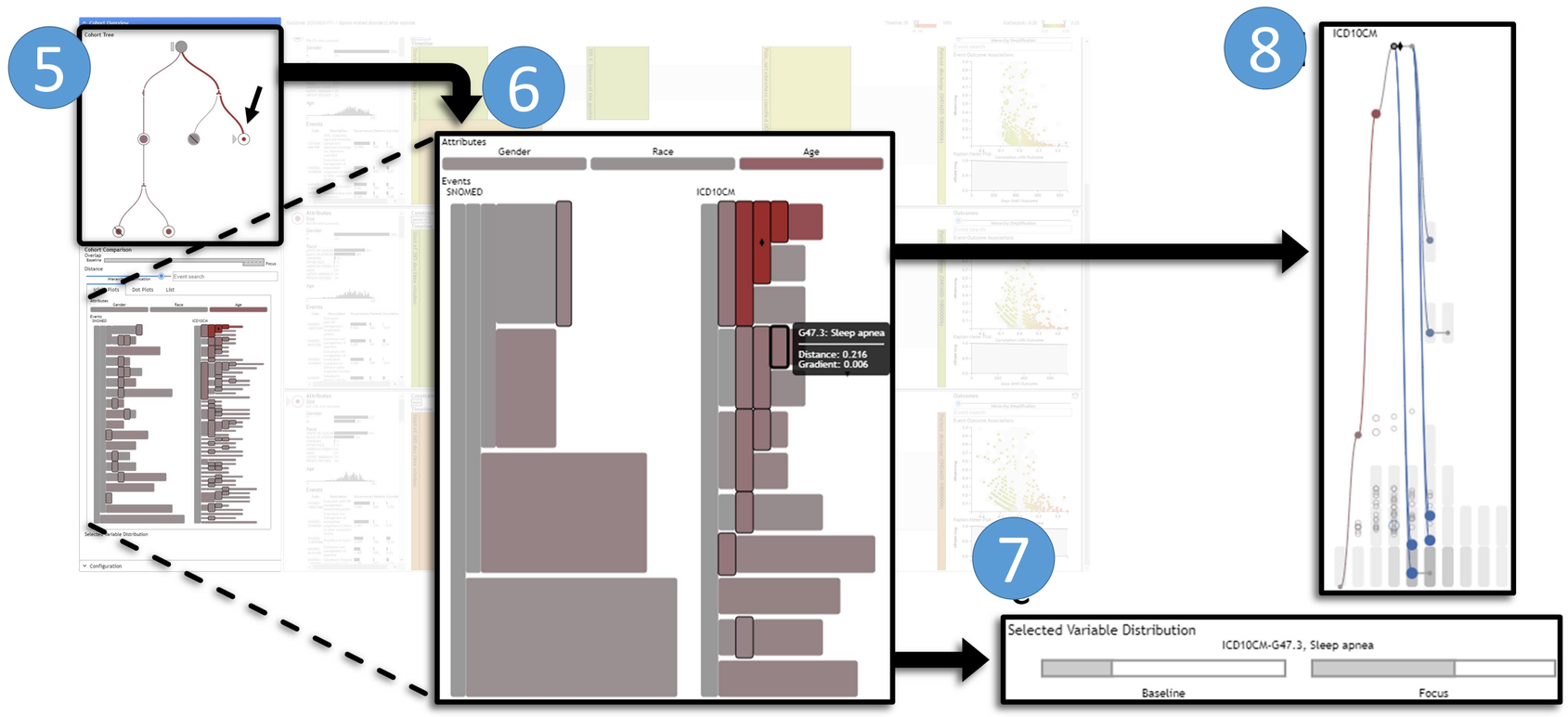
1. ***Summary statistics** for the cohort selected, including the occurrences of each event*
2. ***Timeline view** of relevant events*
3. ***Scatter-plus-focus** view of events for selected cohort; plotting correlation with the outcome against the prevalence. An algorithm selects the hierarchical level of aggregation to use.*
4. ***Hierarchy simplification** slider bar allows dynamic adjustment of aggregation level.*



Cohort Panel

User creates numerous cohorts from the timeline panel. The focused cohort is filtered by Obesity.

5. ***Cohort tree** panel displays the numerous cohorts and path to the current cohort with respect to the baseline*
6. ***Split icicle plot** displays the areas of the dimension hierarchy that contribute most towards selection bias*



7. ***Selected variable distribution visualization** compares distribution of a variable between the selected cohort and baseline cohort*
8. ***Hierarchical dot plot** shows drift, a measure of selection bias, of descendant variables of the constrained variable*

References:

D. Borland, W. Wang, J. Zhang, J. Shrestha, and D. Gotz. Selection bias tracking and detailed subset comparison for high-dimensional data. *IEEE VAST*, 2019.

D. Gotz, J. Zhang, W. Wang, J. Shrestha, and D. Borland. Visual analysis of high-dimensional event sequence data via dynamic hierarchical aggregation. *IEEE VAST*, 2019.