



INTEGO
GROUP

TOPOLOGY-BASED CLINICAL DATA MINING

Identifying Hidden Patterns in Clinical Datasets



EDINBURGH 2017





PART I. INTRODUCTION

Andrey Rekalov, Ph.D., Senior Data Scientist



What challenges industry is facing?

Maximize ROI from Clinical Dataset

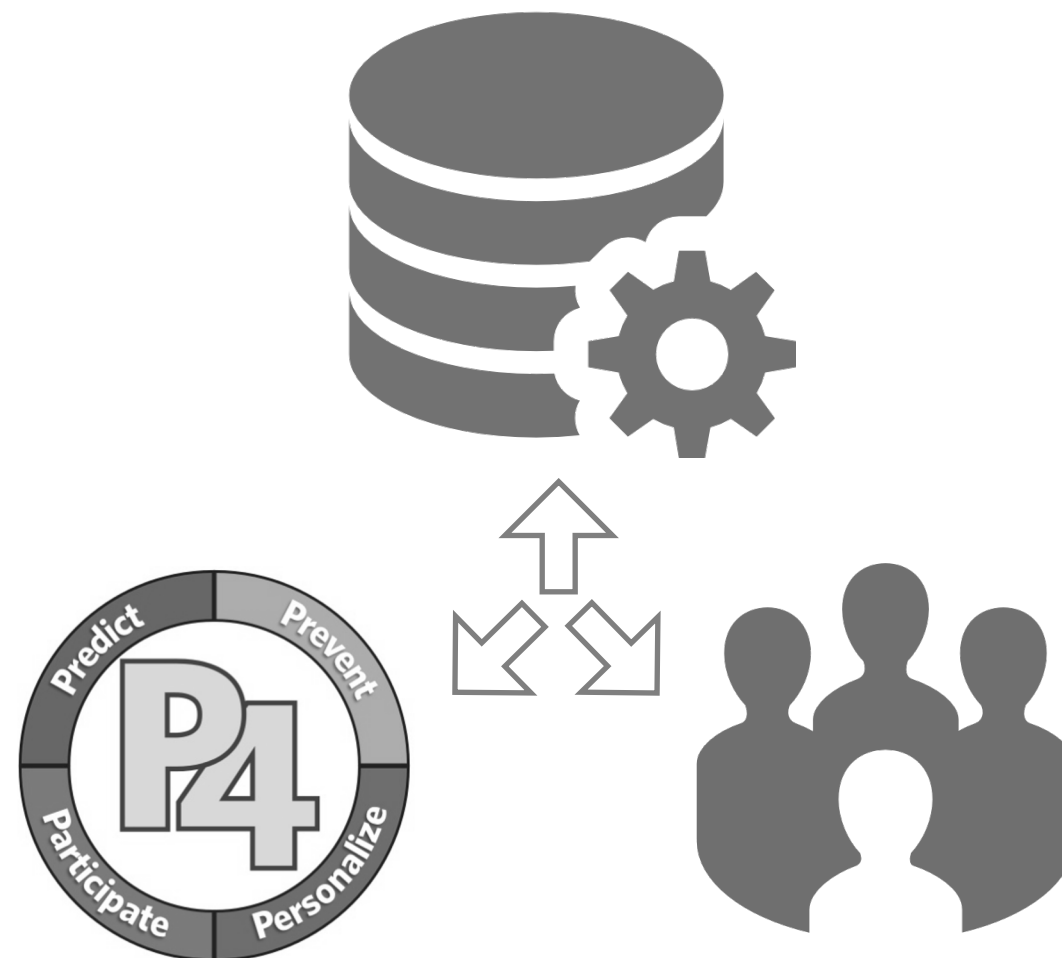
Maximize return on enormous investment made by pharmaceutical companies into clinical study

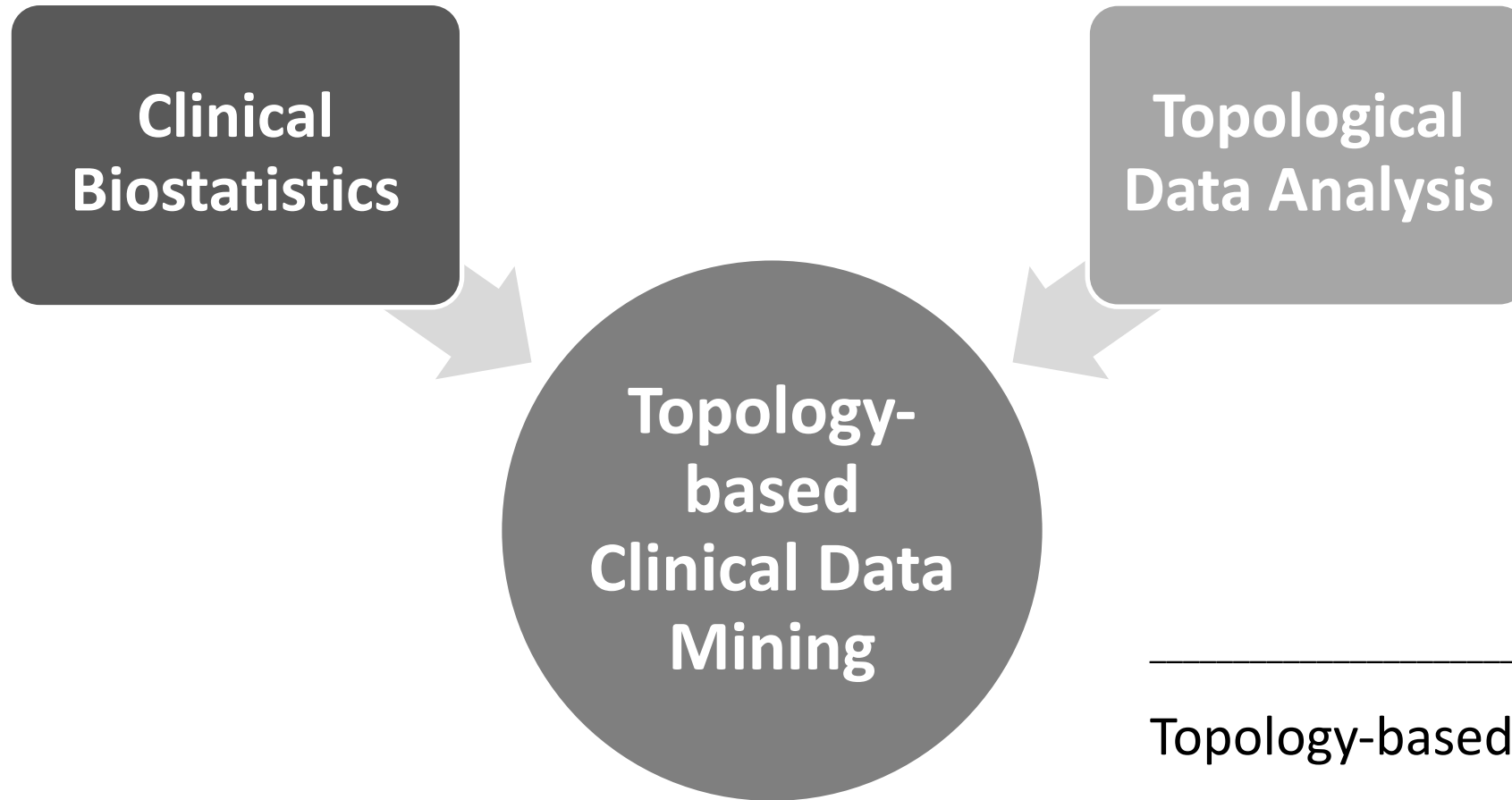
Minimize Research Team's efforts

Cutting cost, time and efforts of the research team by discovering hidden patterns in clinical datasets

Personalized Medicine

Industry needs robust solution for patients segmentations and adverse effect discovery





Topology-based Clinical Data Mining

Application of data mining techniques that involves topological data analysis and biostatistics for extraction, analysis and interpretation of available datasets obtained during clinical trials

What is topological data map?

Each node represents patient(s)

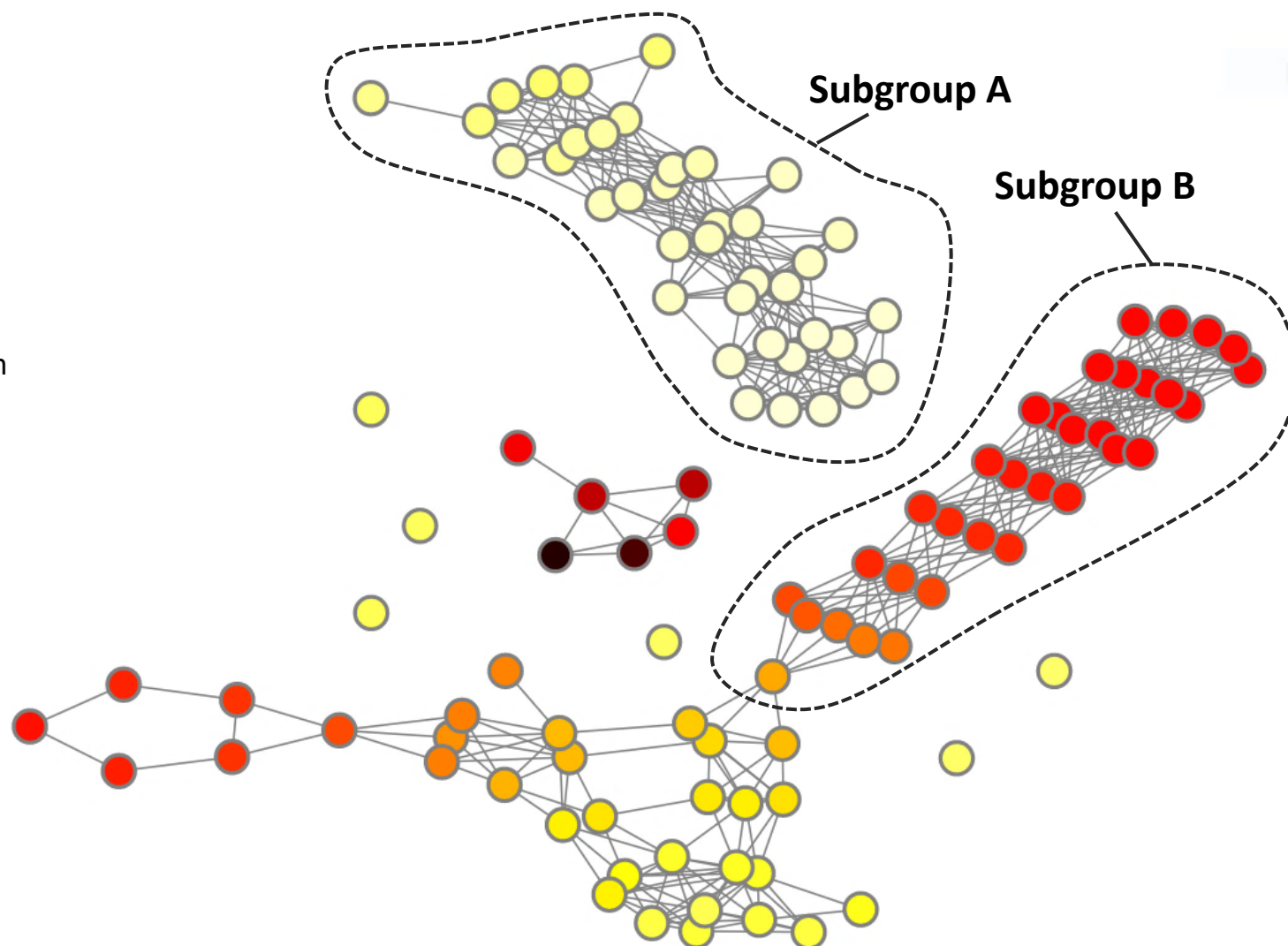
TCDM produces topological data maps, i.e. graphs, where nodes correspond to either individual patients or group of patients within a clinical study

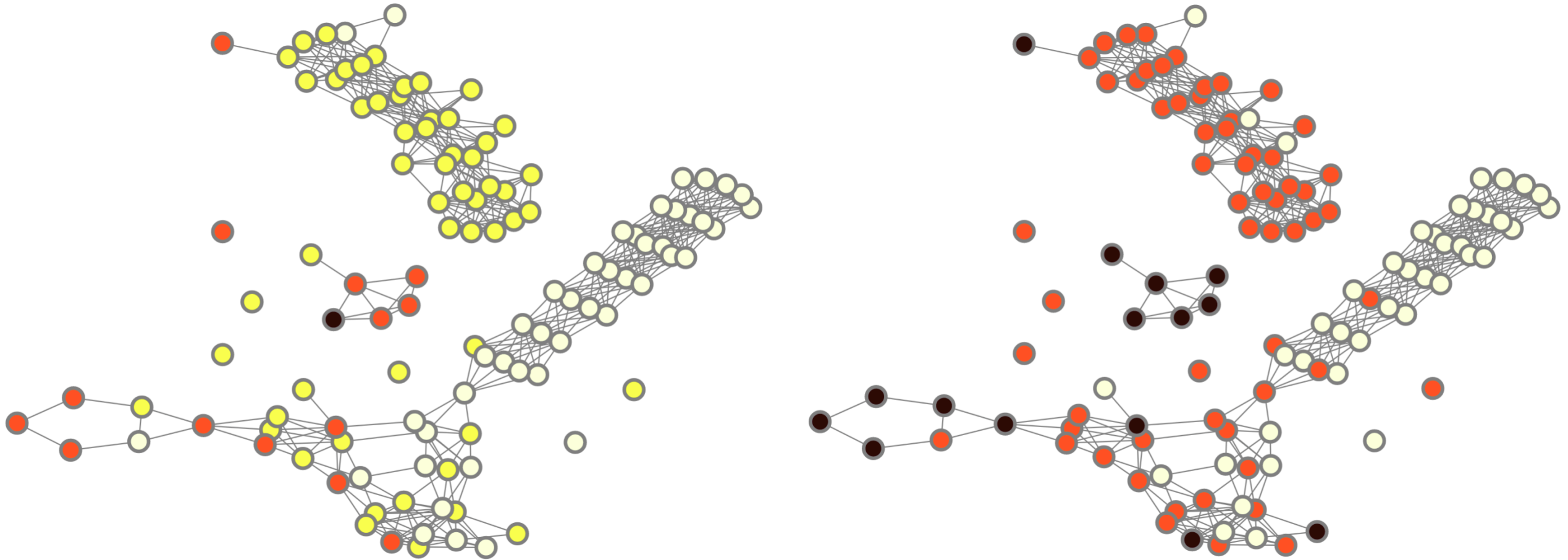
Similar nodes are connected

Two nodes representing similar patients (in terms of a predefined set of clinical outcomes) are connected with an edge

Visual discovery of subgroups

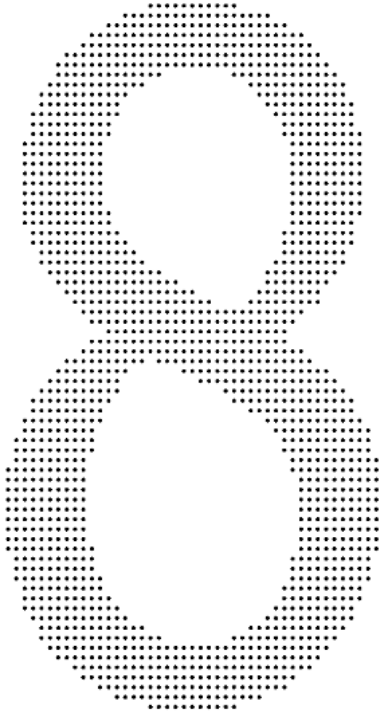
Clusters or "communities" of nodes on a topological data map reflect segmentation of patients which may indicate robust patterns within the data





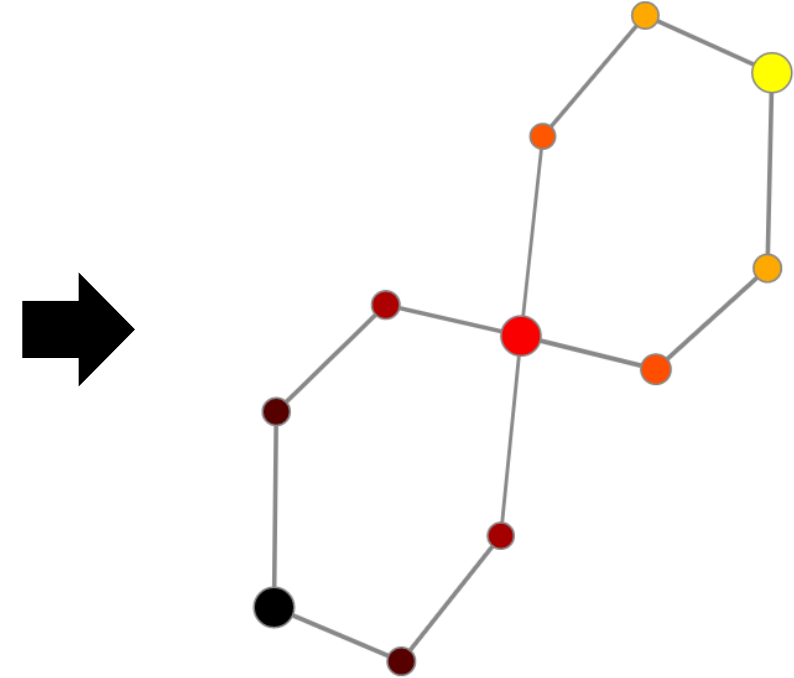
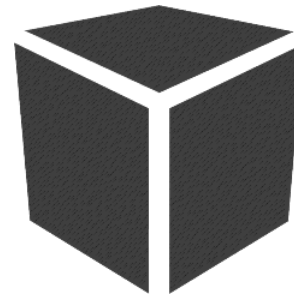
Coloring focused on specific outcomes

Color of the nodes helps highlight emerging patterns in data and identify subgroups of patients related to the distribution of a variable of interest



The digit 8 is a two-dimensional granular dataset consisting of data points with coordinates (x, y)

From dataset to
topological data
map



The topological data map captures the most essential features of the dataset

TCDM can deal with numerical and categorical outcomes

Interrelated biomarkers evaluations

E.g. patients' vital signs or basic metabolic panel's results on a specific day of study

Series of repeated measurements

E.g. weekly hemoglobin levels during chemotherapy in oncological patients

Questionnaire data

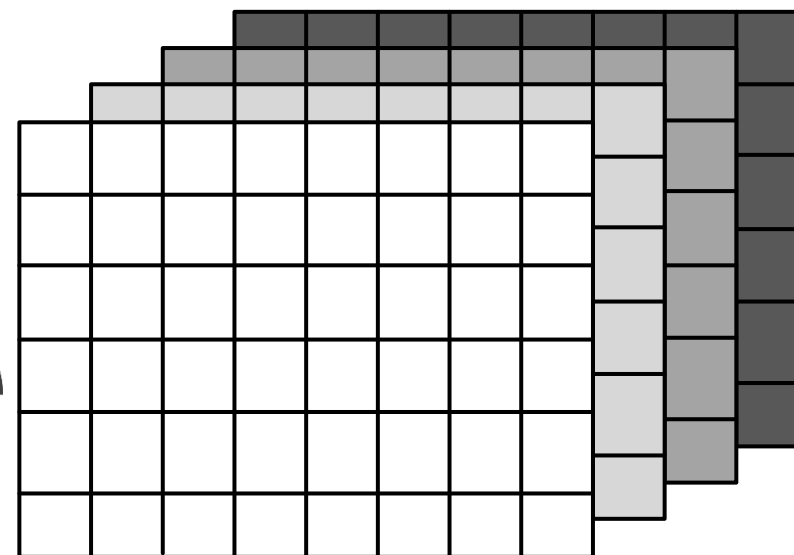
Binary or ordinal responses to the items of a questionnaire, aggregate scores

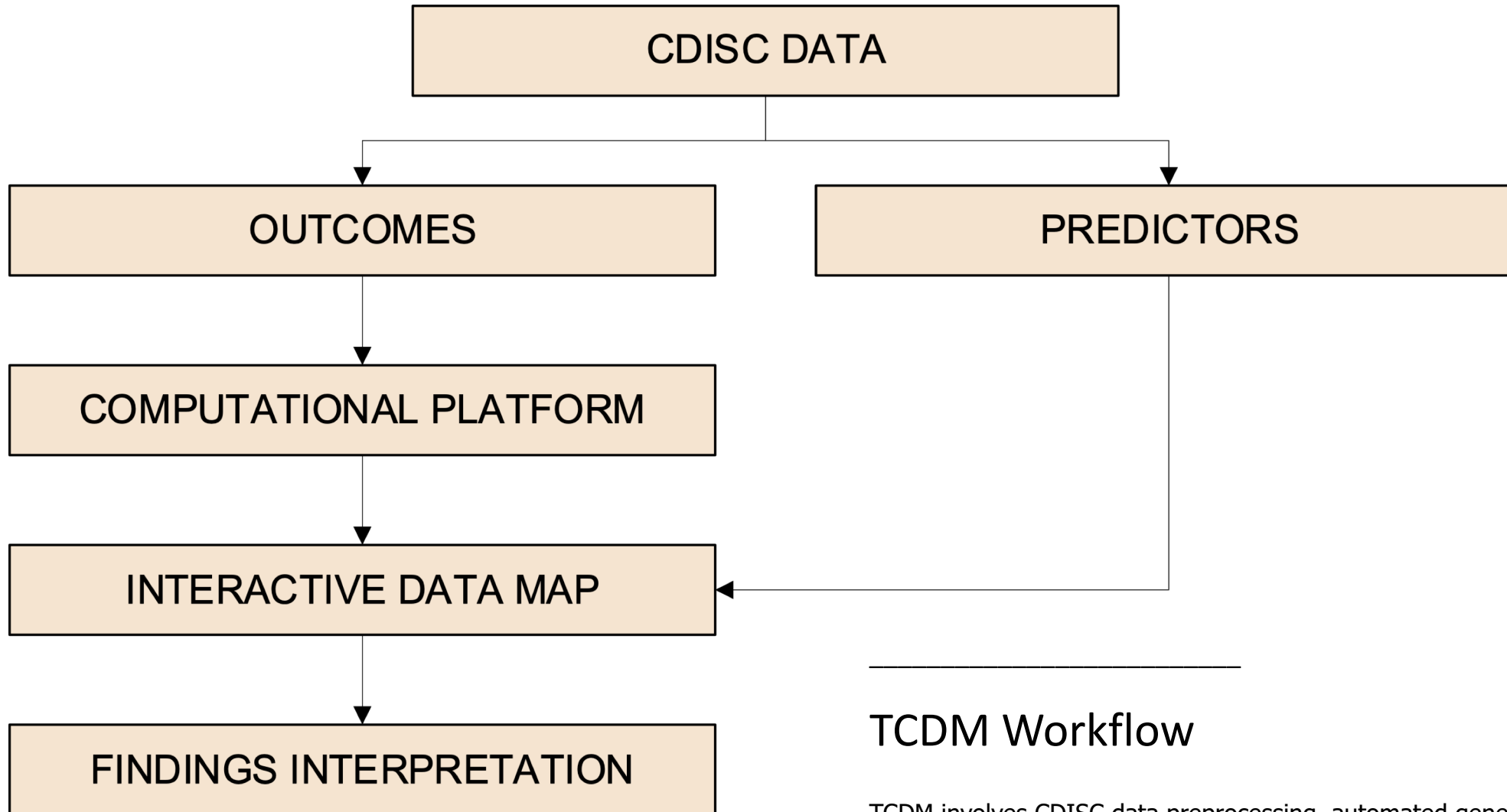


Outcomes



Patients





TCDM Workflow

TCDM involves CDISC data preprocessing, automated generation of topological data maps, visual inspection of interesting features, and statistical analysis of emerging patterns

TCDM versus standard statistical approach



**Standard
statistical
approach**

TCDM



**Outcome variables
are studied
separately**

Allows to analyze
multiple interrelated
outcomes



**Requires certain
assumptions
or a data model**

Assumption-free
and
model-independent



**Hypotheses testing
for pre-specified
subgroups**

Discovery of subgroups
of patients with similar
outcomes



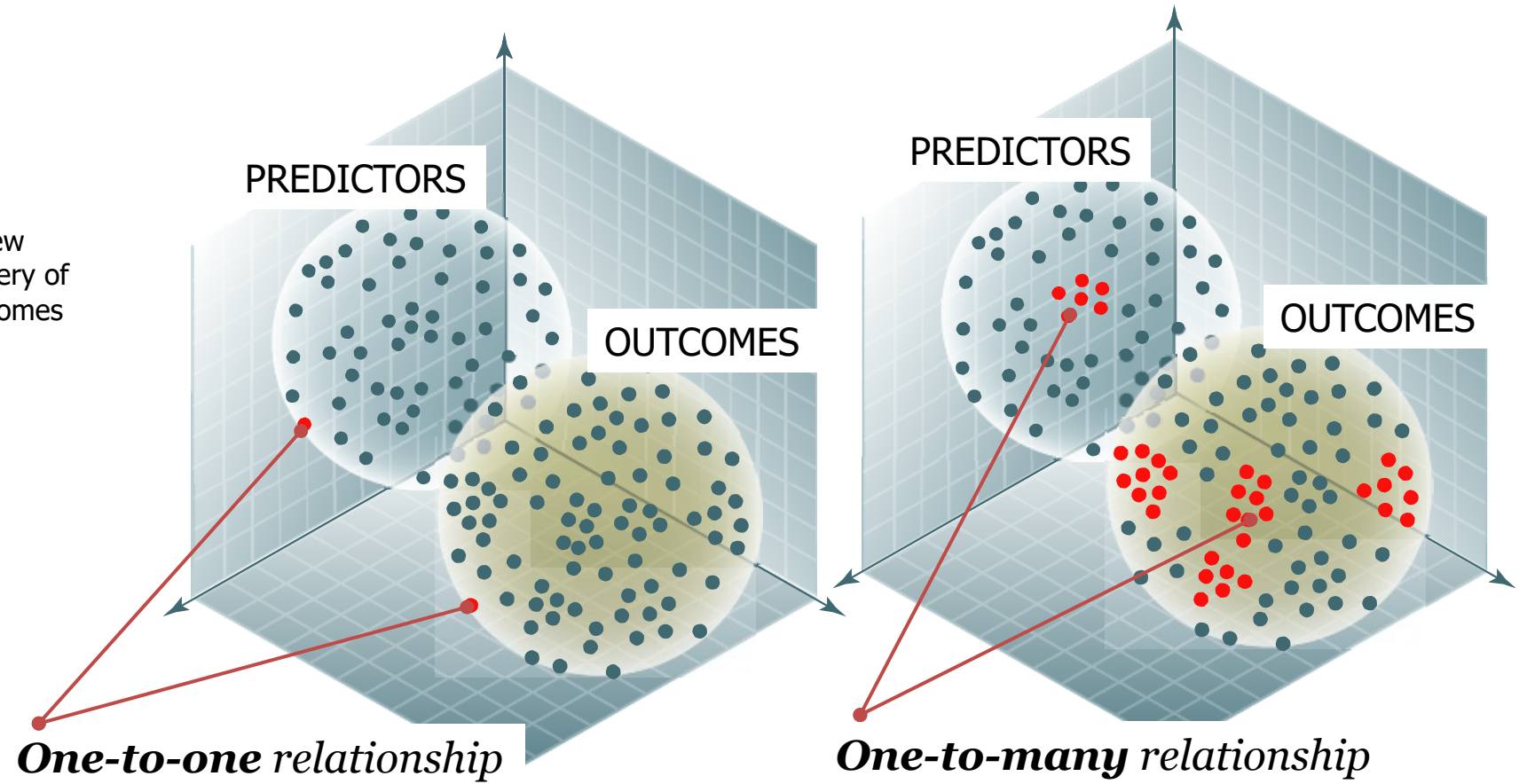
PART II. EXPERIMENT

Iryna Kotenko, Biometrics Group Lead



Univariate vs. Multivariate Analysis

Standard Statistical Analysis usually focuses on relationships between a single outcome and a few covariates. TCDM is designed to facilitate discovery of hidden patterns in multivariate interrelated outcomes



Study for the experiment

A randomized study to test the safety and effectiveness of buprenorphine in the presence of naltrexone for the treatment of cocaine dependence

Primary Outcome Measures

Cocaine use days as measured by self-report, corroborated by thrice-weekly urine drug screens [Time Frame: 30-day evaluation period] 30-day evaluation period is the final 30 days of active medication administration prior to taper; study days 25-54

CTN Protocol ID: CTN-0048

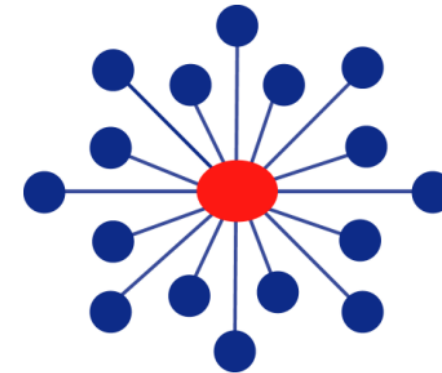
Status: Completed

ClinicalTrials.gov ID: [NCT01402492](https://clinicaltrials.gov/ct/show/NCT01402492)

Link: <https://www.clinicaltrials.gov/ct/show/NCT01402492?order=1>

De-Identification:

<https://datashare.nida.nih.gov/sites/default/files/studydocs/272/CTN0048%20Deidentification%20Notes.pdf>



NIDA CTN Protocol 0048

**Cocaine Use Reduction with
Buprenorphine
(CURB)**

Lead Investigator: Walter Ling, M.D.

March 6, 2012

Version 4.0

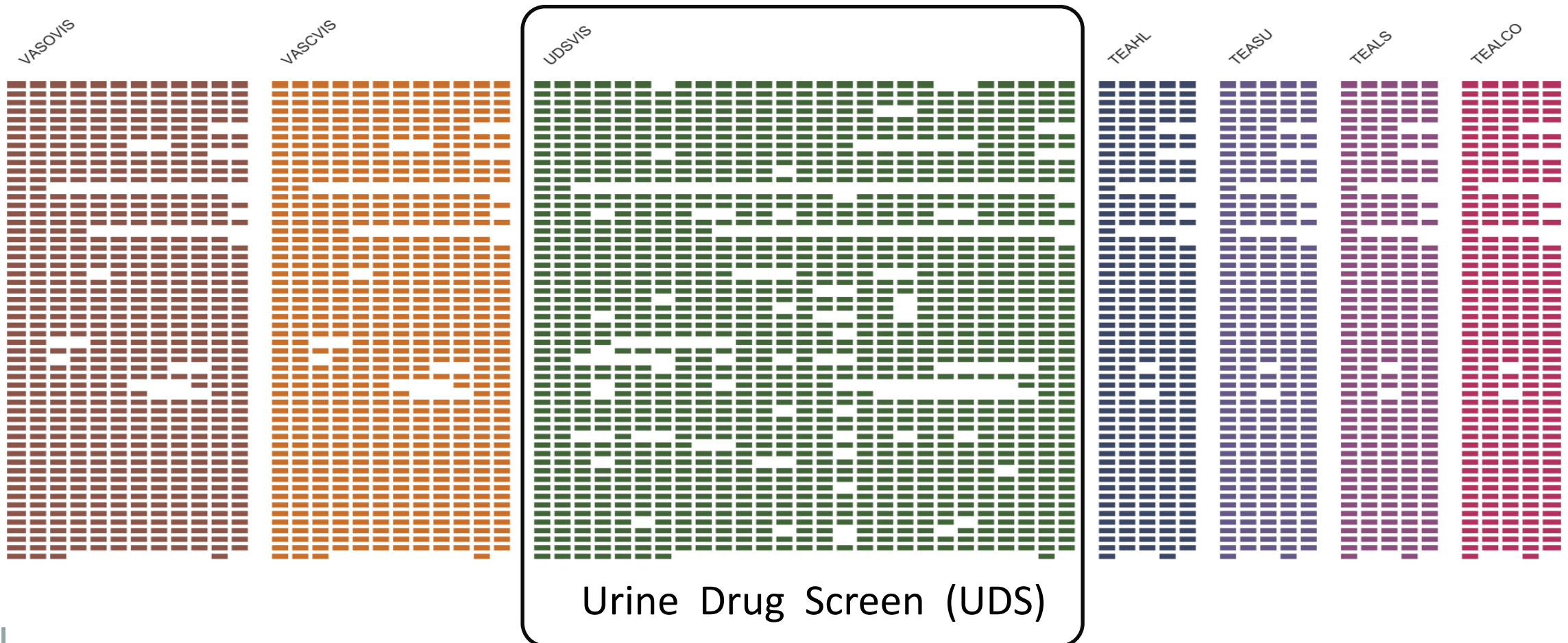
Data explanation and pre-processing using SAS®



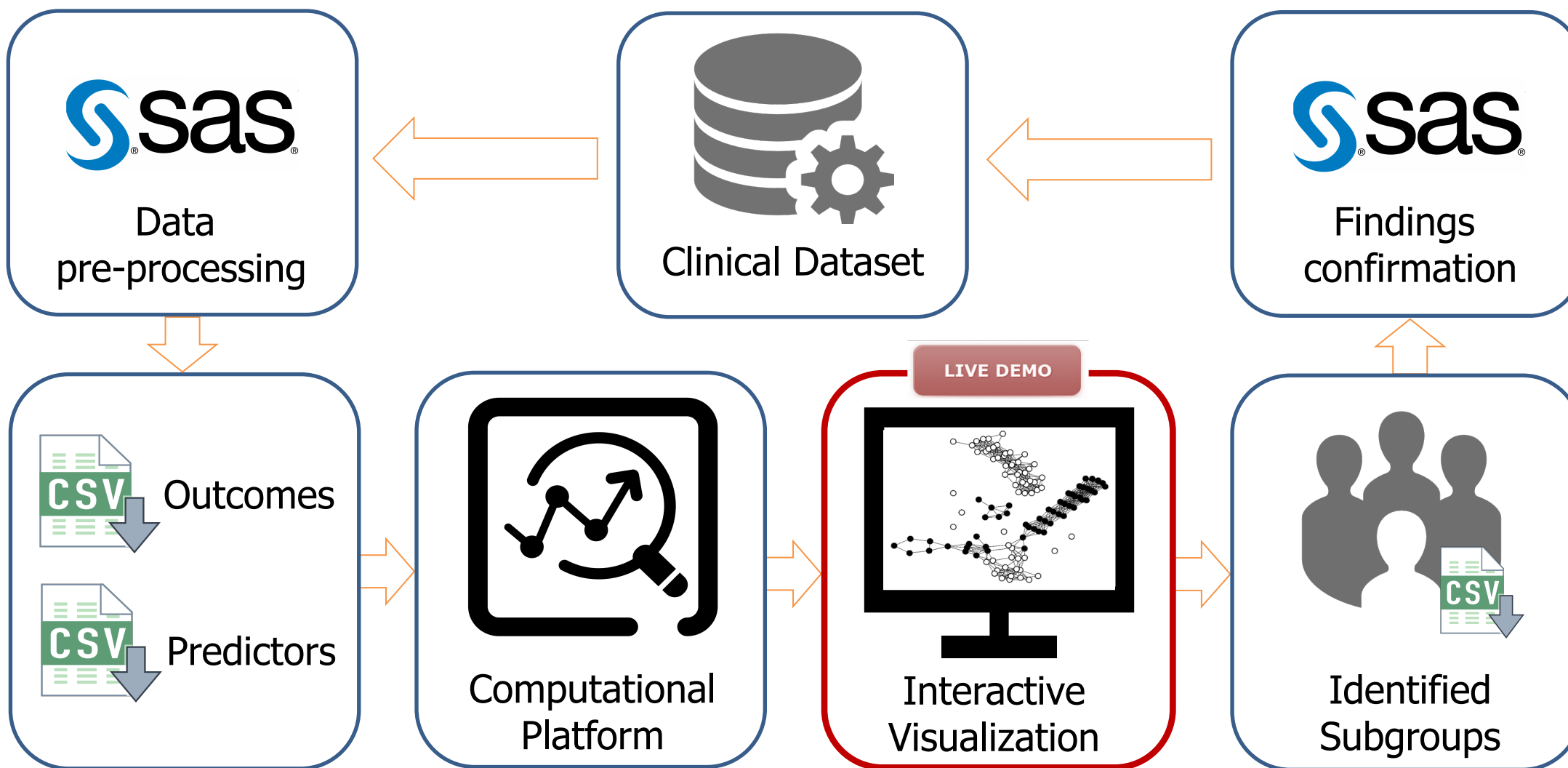
	Protocol	CTP	Participant ID	Visit Number	Urine collection date	Urine 1 collection time	Urine 1 temp in range	Urine 1 adulterated	Urine 1 benzodiazepines	Urine 1 amphetamine	Urine 1 THC marijuana	Urine 1 methamphetamine
1	0048	938	81653	102	3	12:10	1	0	0	0	0	0
2	0048	938	81653	103	5	09:04	1	0	0	0	0	0
3	0048	938	81653	201	7	08:59	1	0	0	0	0	0
4	0048	938	81653	202	10	13:46	1	0	0	0	0	0
5	0048	938	81653	203	12	08:29	1	0	0	0	0	0

	usubjid	Number of substances indicated in urine at visit 00	Number of substances indicated in urine at visit 101	Number of substances indicated in urine at visit 102	Number of substances indicated in urine at visit 103	Number of substances indicated in urine at visit 201	Number of substances indicated in urine at visit 202	Number of substances indicated in urine at visit 301	Number of substances indicated in urine at visit 302	Number of substances indicated in urine at visit 303
1	NIDA-0048-091-108859	1	1	1	1	0	0	1	1	2
2	NIDA-0048-091-135907	2	2	2	1	2	2	1	2	2
3	NIDA-0048-091-174303	1	1	1	1	2	1	1	1	1
4	NIDA-0048-091-345115	1	1	1	1	1	1	1	1	1
5	NIDA-0048-091-502273	1	1	1	1	1	1	0	1	1

Outcomes grouped by type

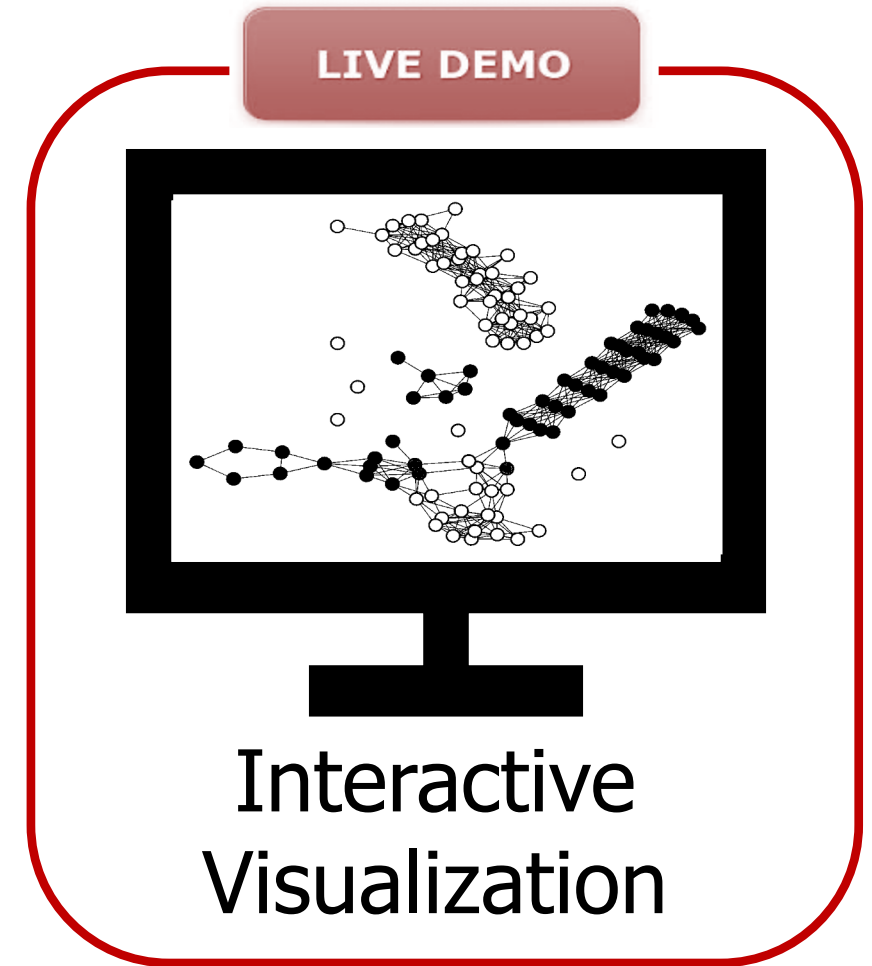


EXPERIMENT WORKFLOW



EXPERIMENT RESULTS

LIVE DEMO



Experiment wrap up

Subgroup A

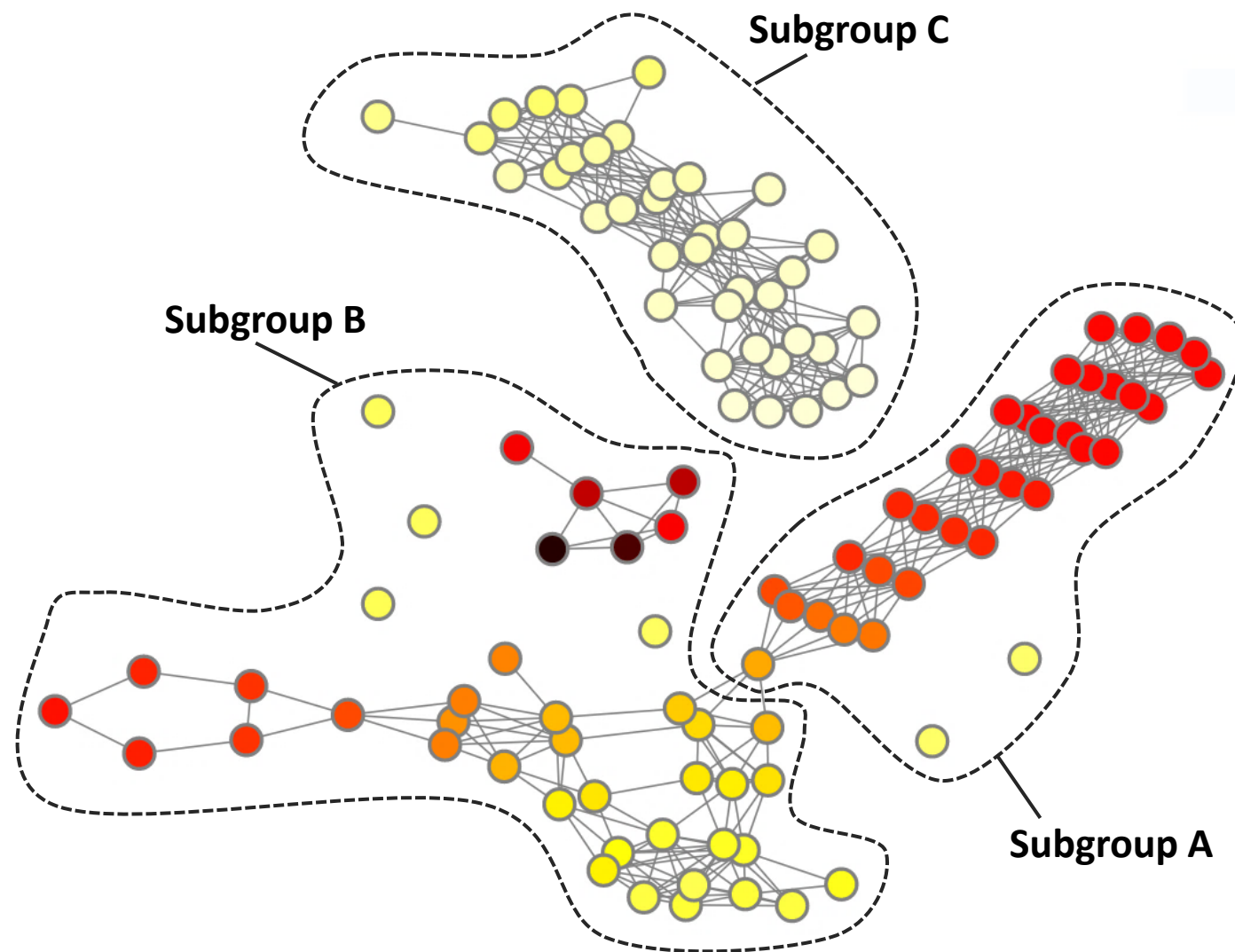
On average, the patients of this group were older than those of Subgroup B and Subgroup C

Subgroup B

Subgroup B contained more patients who had a history of benzodiazepines abuse in comparison to the patients in Subgroup C

Subgroup C

The patients of Subgroup C exhibited a lower depression score throughout the study in comparison to those in Subgroup B. They also had a history of physical and neurotic disorders





THANK YOU

Intego Group, LLC
555 Winderley Place, Ste. 129,
Maitland, FL 32751
Phone: +1 (407) 641-4730
sergey.glushakov@intego-group.com
www.intego-group.com

