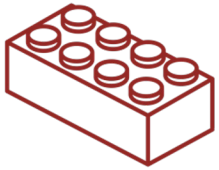
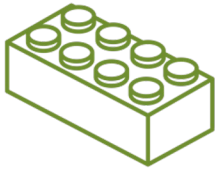
Characterization

Data are Like Lego Bricks for Phenotypng

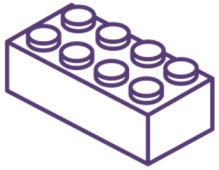
Conditions



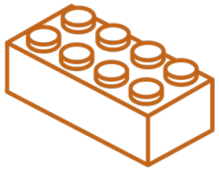
Drugs



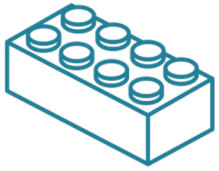
Procedures



Measurements



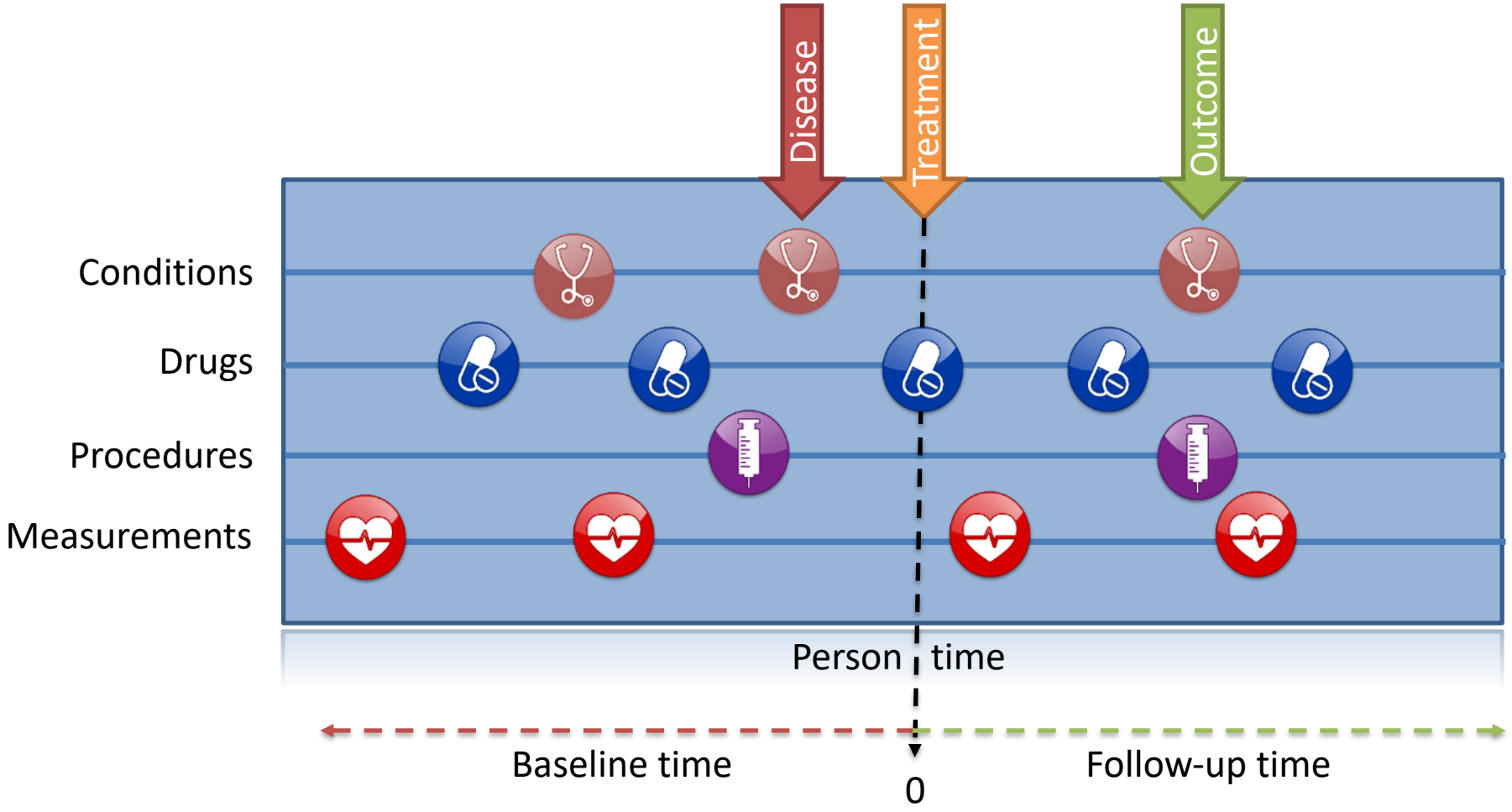
Observations



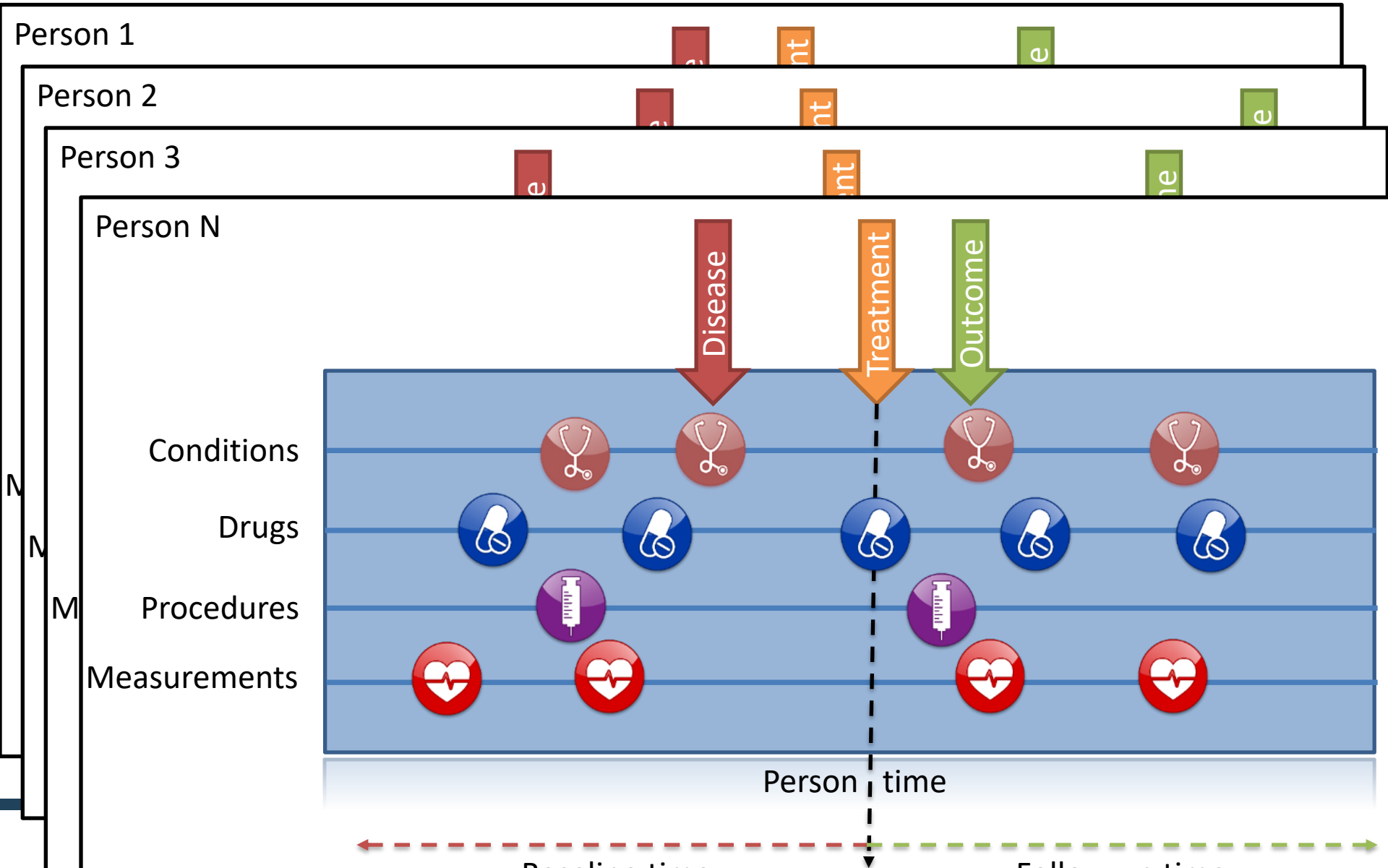
Visits

Why bother with characterization?

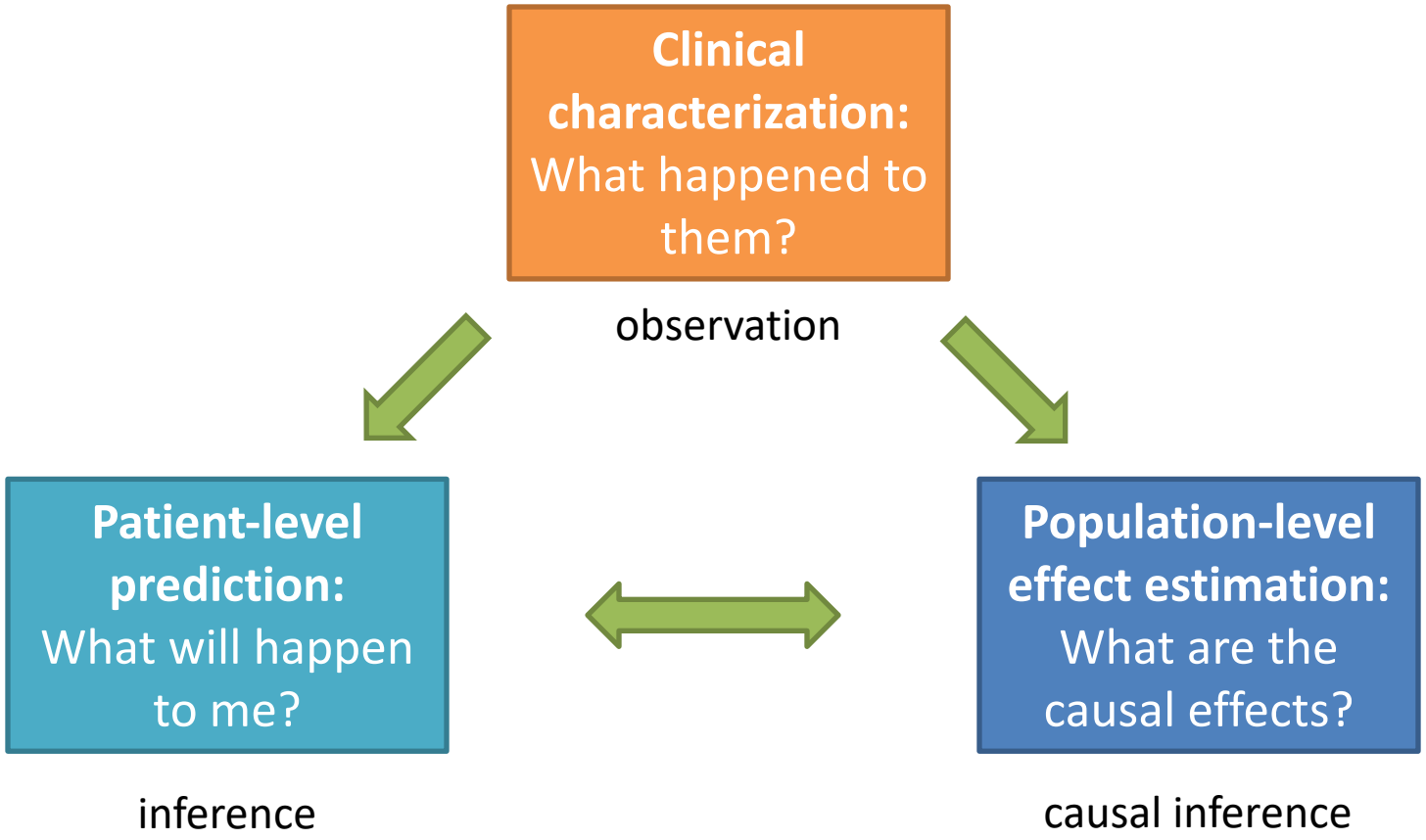
A caricature of the patient journey



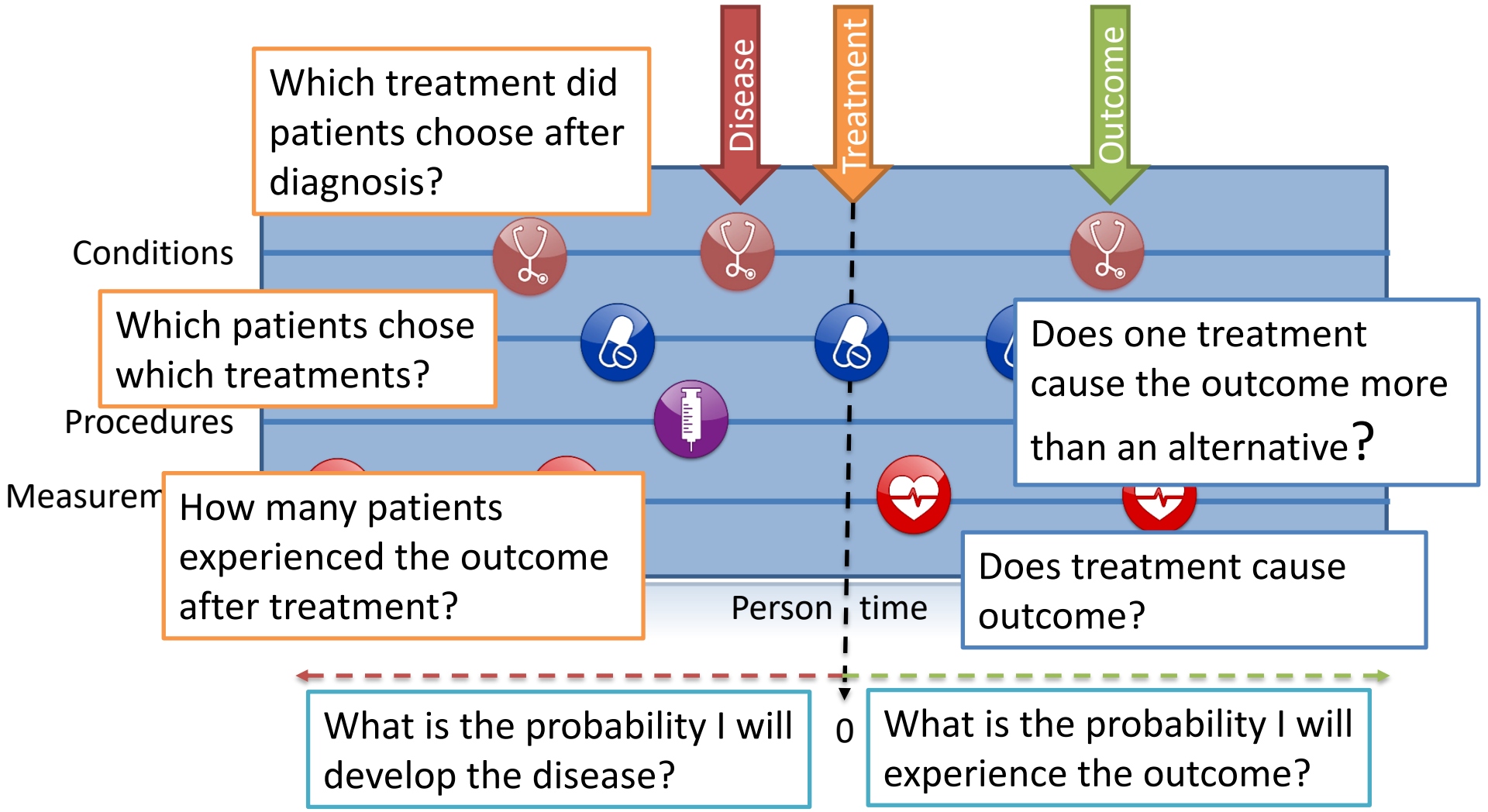
Each observational database is just an (incomplete) compilation of patient journeys



Complementary evidence to inform the patient journey



Questions asked across the patient journey



HADES Spotlight: FeatureExtraction

An R package for generating features (covariates) for a cohort using data in the Common Data Model

Best supporting package (IMHO)

Using FeatureExtraction

Martijn J. Schuemie

2021-07-26

Contents

1 Introduction

The `FeatureExtraction` package can be used to create features for a cohort, using the information stored in the Common Data Model. A cohort is defined a set of persons who satisfy one or more inclusion criteria for a duration of time. Features can for example be diagnoses observed prior to entering the cohort. Some people might also refer to such features as 'baseline characteristics', or to features in general as 'covariates', and we will use those terms interchangeably throughout this vignette.

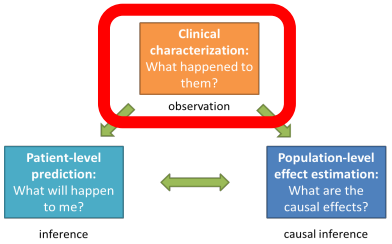
This vignette describes how features can be constructed using the default covariate definitions embedded in the package. Although these definitions allow quite some customization through predefined parameters, it is possible that someone needs more customization. In this case, the reader is referred to the other vignettes included in this package that deal with constructing completely custom covariates.

This vignette will first describe how to specify which features to construct. In many situations, for example when using `FeatureExtraction` as part of another package such as `CohortMethod` or `PatientLevelPrediction`, that is all one needs to know about the `FeatureExtraction` package, as the actual calling of the package is done by the other package. However, it is also possible to use this package on its own, for example to create a descriptive characterization of a cohort to include in a paper.

1

https://github.com/ohdsi/featureextraction

OHDSI Characterization Framework



- Define a characterization study in terms of:
- Target cohorts (T): those to characterize
- Subgroup cohorts (S): those to use as subgroups of the target cohort(s)
- Feature cohorts (F): cohorts used to construct features (outcomes) for characterization
- Time at risk windows: Define windows of time to characterize all features (F) and concepts

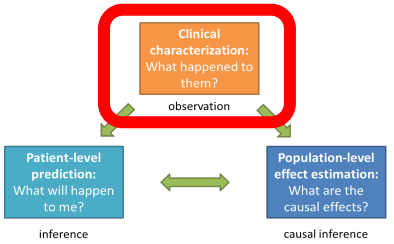
OHDSI's definition of 'cohort'

Cohort = a set of persons who satisfy one or more inclusion criteria for a duration of time

- One person may belong to multiple cohorts
- One person may belong to the same cohort at multiple different time periods
- One person may not belong to the same cohort multiple times during the same period of time
- One cohort may have zero or more members
- A codeset is NOT a cohort…
  …logic for how to use the codeset in a criteria is required

Cohort = Phenotype for a duration of time

OHDSI Characterization Framework



- Target cohort: who do you want to study?
- Stratification (pre-index): what subgroups do you want to study?
- Features of interest: what attributes do you want to look at and describe differences in?
- Time-at-risk: what windows of time do you want to describe features in?

Cohort Diagnostics

Descriptive statistics that provide insight on the performance of multiple cohort definitions when applied across data sources

Cohort Counts
"Magnitude of difference"

Incidence Rate
"Baseline expectation"

Time Distribution
"Before, During or After"

Index Breakdown
"Triggered Entry"

Visit Context
"Patient care setting"

Cohort Overlap
"Common vs. Different"

Temporal Characterization
"Before, On the day of, After"

Incidence Analysis

Cohort Pathways

OHDSI in action: Clinical characterization

PNAS

COLLOQUIUM PAPER

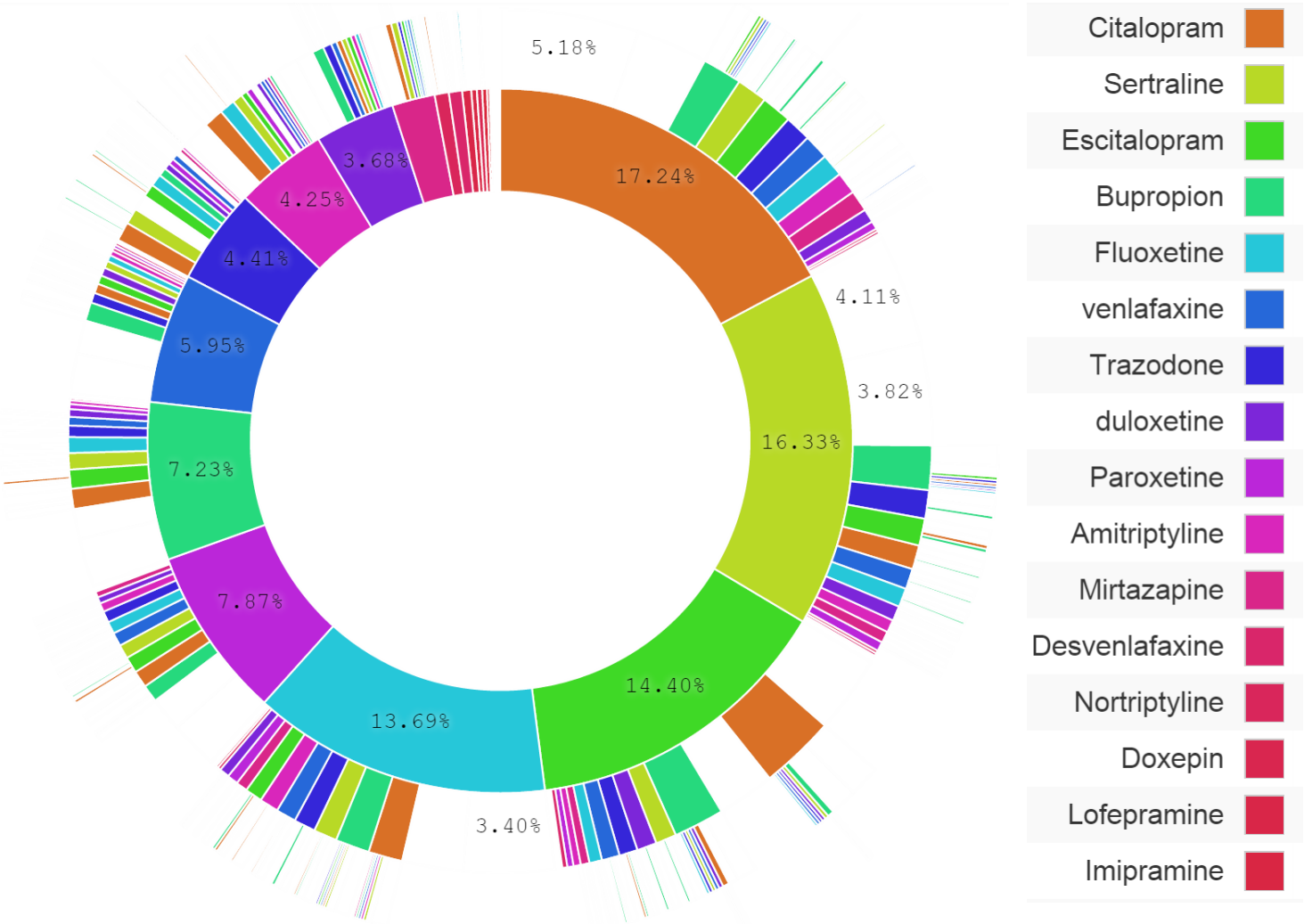
Characterizing treatment pathways at scale using the OHDSI network

George Hripcsak[a,b,c,1], Patrick B. Ryan[c,d], Jon D. Duke[c,e], Nigam H. Shah[c,f], Rae Woong Park[c,g], Vojtech Huser[c,h], Marc A. Suchard[c,i,j,k], Martijn J. Schuemie[c,d], Frank J. DeFalco[c,d], Adler Perotte[a,c], Juan M. Banda[c,f], Christian G. Reich[c,l], Lisa M. Schilling[c,m], Michael E. Matheny[c,n,o], Daniella Meeker[c,p,q], Nicole Pratt[c,r], and David Madigan[c,s]

[a]Department of Biomedical Informatics, Columbia University Medical Center, New York, NY 10032; [b]Medical Informatics Services, NewYork-Presbyterian Hospital, New York, NY 10032; [c]Observational Health Data Sciences and Informatics, New York, NY 10032; [d]Epidemiology Analytics, Janssen Research and Development, Titusville, NJ 08560; [e]Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN 46205; [f]Center for Biomedical Informatics Research, Stanford University, CA 94305; [g]Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, South Korea, 443-380; [h]Lister Hill National Center for Biomedical Communications (National Library of Medicine), National Institutes of Health, Bethesda, MD 20894; [i]Department of Biomathematics, University of California, Los Angeles, CA 90095; [j]Department of Biostatistics, University of California, Los Angeles, CA 90095; [k]Department of Human Genetics, University of California, Los Angeles, CA 90095; [l]Real World Evidence Solutions, IMS Health, Burlington, MA 01809; [m]Department of Medicine, University of Colorado School of Medicine, Aurora, CO 80045; [n]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37212; [o]Geriatric Research, Education and Clinical Center, VA Tennessee Valley Healthcare System, Nashville, TN 37212; [p]Department of Preventive Medicine, University of Southern California, Los Angeles, CA 90089; [q]Department of Pediatrics, University of Southern California, Los Angeles, CA 90089; [r]Division of Health Sciences, University of South Australia, Adelaide, SA, Australia 5001; and [s]Department of Statistics, Columbia University, New York, NY 10027

Edited by Richard M. Shiffrin, Indiana University, Bloomington, IN, and approved April 5, 2016 (received for review June 14, 2015)

How are patients with major depressive disorder *ACTUALLY* treated?



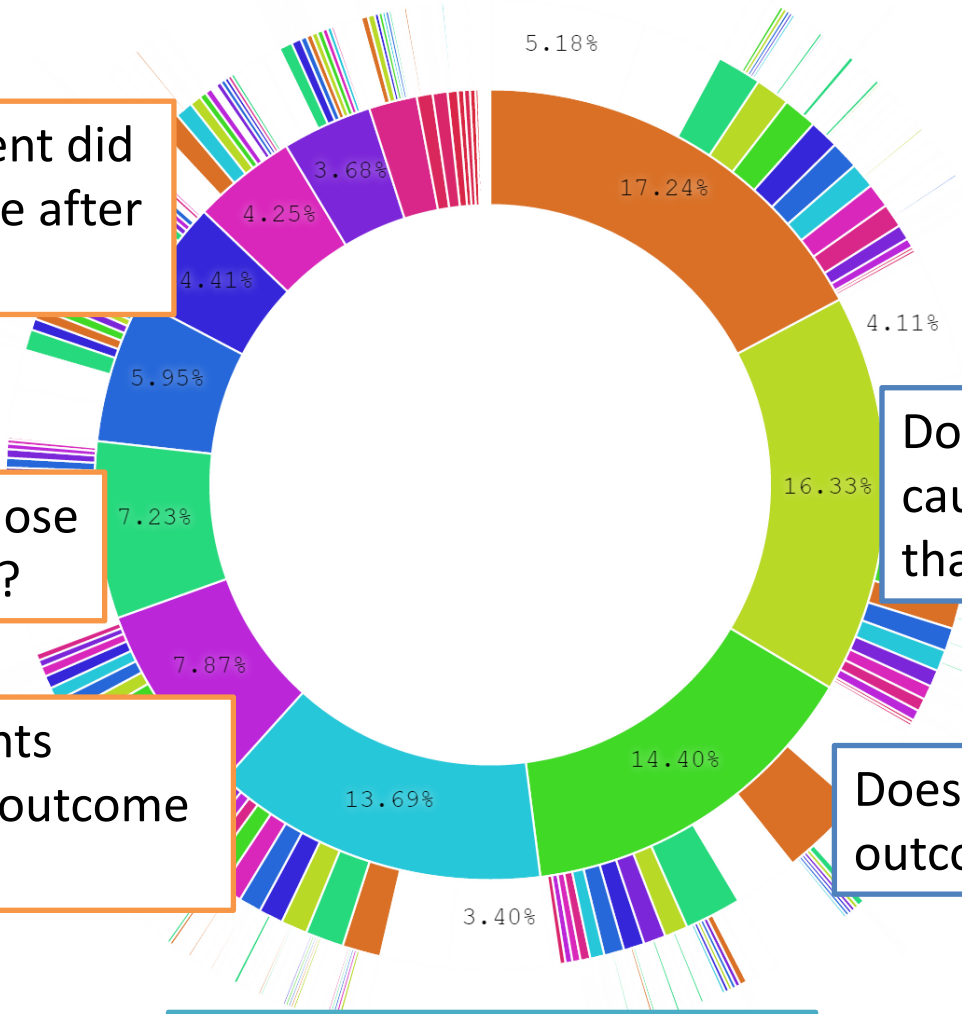
Hripcsak et al, PNAS, 2016

How are patients with major depressive disorder *ACTUALLY* treated?

- Substantial variation in treatment practice across data sources, health systems, geographies, and over time
- Consistent heterogeneity in treatment choice as no source showed one preferred first-line treatment
- 11% of depressed patients followed a treatment pathway that was shared with no one else in any of the databases

Hripcsak et al, PNAS, 2016

What questions does this answer?
What question does it prompt to ask?



Citalopram
Sertraline
Escitalopram
Bupropion
Fluoxetine
venlafaxine
Mirtazapine
Desvenlafaxine
Lofepramine
Imipramine

Which treatment did patients choose after diagnosis?

Which patients chose which treatments?

How many patients experienced the outcome after treatment?

Does one treatment cause the outcome more than an alternative?

Does treatment cause outcome?

What is the probability I will experience the outcome?

Hripcsak et al, PNAS, 2016

Demo #1: Incidence of myocardial infarction among new users of lisinopril in 'on treatment' time-at-risk

Let’s play with ATLAS!

Exercise: Incidence of angioedema among new users of lisinopril in 'on treatment' time-at-risk

Extra credit #1: Incidence of myocardial infarction among new users of lisinopril in 'on treatment' time-at-risk, within subpopulations of interest (age<18, women, Black)

"You'll never walk alone on your OHDSI Journey."

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Questions?

kostka@ohdsi.org

Join the Journey

http://ohdsi.org