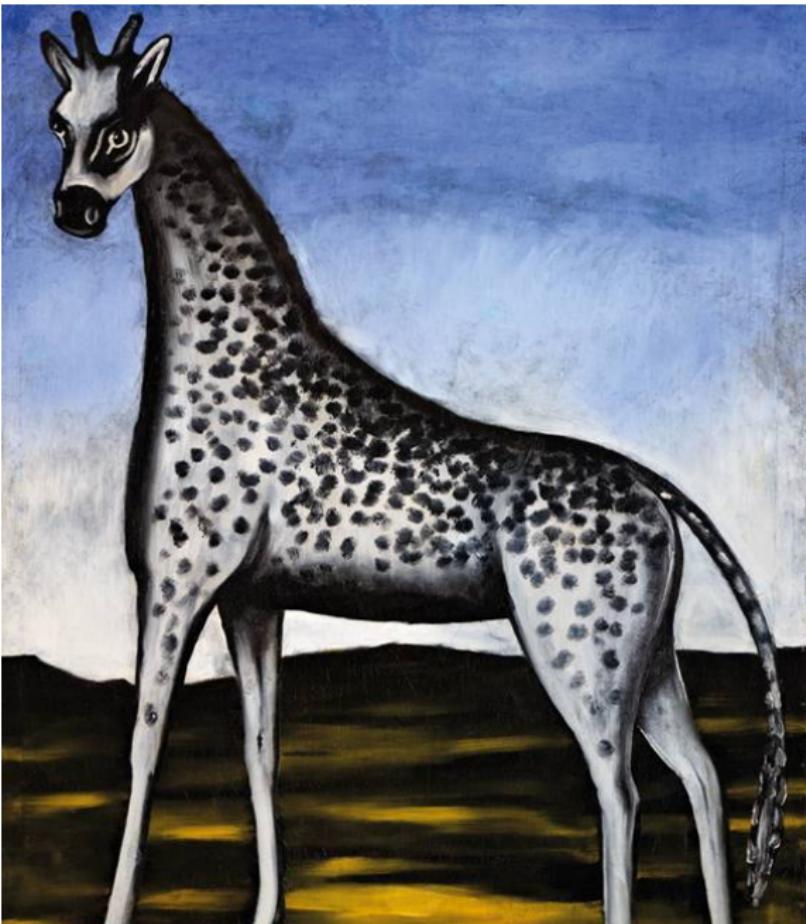


# Reddie WP4

Christian Torp-Pedersen

May 21, 2023

From register data to covariates, treatment, outcomes



### Therapies in diabetes using four real-world databases

- ▶ W4.1: Establish a common analytical platform using longitudinal targeted maximum likelihood estimation (LTMLE) that can be used across the four databases
- ▶ W4.2: Assess how models derived from RCT data perform within RWD using a matched cohort:
- ▶ W4.3: Assess how models derived from RCT data perform within RWD from all that received the intervention
- ▶ W4.4: Integrate results from the four databases and perform meta-analysis.

## LTMLE general constraint

- ▶ LTMLE requires that follow-up is divided in a number of equally length time intervals
- ▶ The positivity assumption of causal inference requires that the probability of exposure to any covariate is above zero and below one in each interval
- ▶ Therefore the subdivision of time needs to be a compromise
- ▶ In our analyses so far we have used 6 month intervals. When we discuss how to define exposure to treatments the choice of interval is important

## How to define type 2 diabetes

- ▶ Classification in a register
- ▶ A diagnosis in a register
- ▶ Use of hypoglycemic therapy
  - ▶ Anyone starting metformin
  - ▶ Women below 40 using only metformin likely have polycystic ovarian syndrome
  - ▶ Pregnant women only using therapy during pregnancy have pregnancy related diabetes
  - ▶ People starting therapy with insulin are less likely to have type 2 diabetes

## Covariates

Age - Sex - Body mass index - Income - Degree of urbanization of home address (rural, suburban, or urban) (DEGURBA) - Education, ISCED class - Duration of Diabetes - Duration and type of first line hypoglycemic therapy - Duration and type of second line hypoglycemic therapy - HbA1c - Smoking yes/no - History of Ischemic heart disease - History of Heart Failure - History of stroke - History of Peripheral artery disease - History of renal disease - History of diabetic eye disease - History of chronic obstructive pulmonary disease - History of hypertension - Therapy with aldosterone receptor antagonists, beta blocker, angiotensin converting enzyme inhibitor/angiotensin receptor-2 blocker, loop diuretics, calcium channel blockers, thiazide, acetylsalicylic acid, statins Blood pressure Urine albumine creatinine ratio - Glomerular filtration rate based on creatinine (eGFR)

## Time dependent variables

Glucose measurements - HbA1c levels - Hypoglycemic treatment categories - Atherosclerotic cardiovascular disease (myocardial infarction, stroke, peripheral cardiovascular disease) - Renal disease - Atrial fibrillation - Hypertension - Heart Failure - Chronic obstructive lung disease - Statin treatment - Aspirin or other anticoagulation treatment

# Good Bad and Ugly covariates



## Good covariates

- ▶ Age and Sex
- ▶ Diagnoses that are common and dominate a clinical situation - examples are atherosclerotic disease, stroke and chronic pulmonary disease
- ▶ Laboratory values that are taken in a systematic manner
- ▶ Individually based information on education and income

## Bad covariates

- ▶ Diagnoses that reflect less common diseases.
- ▶ Diagnoses where multiple condition converge - e.g. Heart failure and Renal disease
- ▶ Laboratory values that reflect a mixture of the actual value and the reason for obtaining it - perhaps HbA1c
- ▶ Systematic missingness

## Ugly covariates

- ▶ Conditions that are overlooked or falsely reported - smoking, alcohol
- ▶ Pharmaceutical treatments - all reflect a mixture of the indication and the signal of being willing to receive treatment
- ▶ Surgical treatments - by pass operation

## Our selection!

- ▶ Variables that are decently trustworthy
- ▶ Differences between the four countries should be logical
- ▶ Avoid treatments?
- ▶ Avoid missingness

## Outcome

- ▶ Major Adverse Cardiovascular Events (MACE)
- ▶ AMI or Stroke
- ▶ AMI or Stroke or Heart Failure
- ▶ Glycemic result - requires systematic collection

## Practical data preparation for LTMLE

- ▶ A set of equally length time periods
- ▶ Anodes - Variables that are treatment value for each period
- ▶ Lnodes - Covariates for each period - also includes competing risk
- ▶ Ynodes - Outcome in each period
- ▶ Cnodes - Censoring for each period

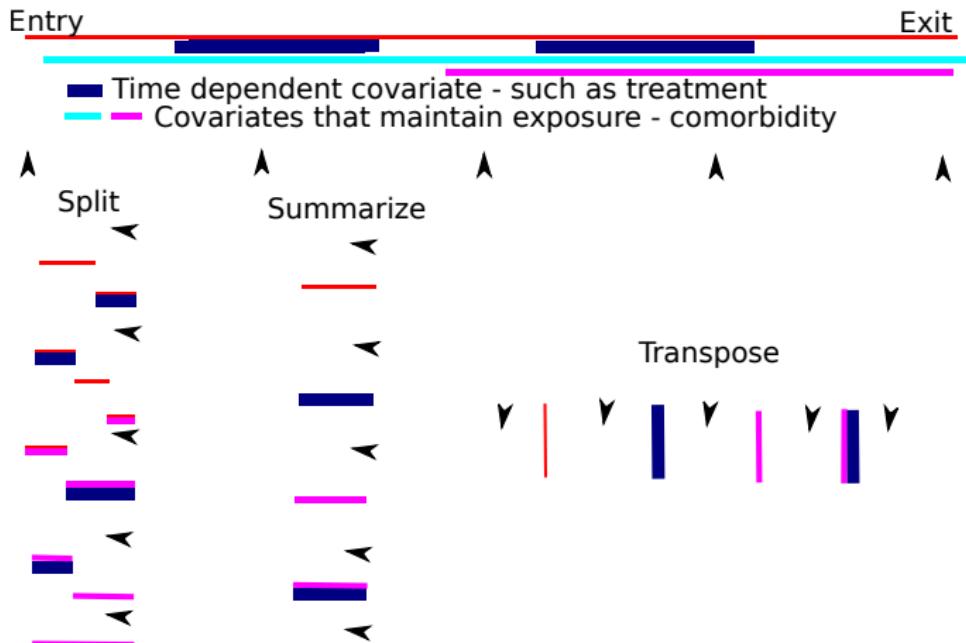
The variables have to be specified in a precise order and there are rules to obey!

# Practical LTMLE

- ▶ Today we present the path from a set of data tables to a final dataset that can enter an analysis
- ▶ Tomorrow comes how to make a proper LTMLE analysis with the data
- ▶ The example I go through today may or may not reflect data in four countries, but from step 2 the paths could converge
- ▶ There is no rule as to how to achieve the final analysis data, but what is shown today works

## Build four datasets

- ▶ A dataset with entrydate and final exit date for each individual
  - If e.g. 10 periods of 6 months is the analysis, then  $10 * 180$  days is added to the entry date to obtain the exit date
- ▶ A dataset with fixed baseline variables
- ▶ A dataset with time dependent variables that can change only once - including outcomes, censoring and competing risk variables
- ▶ A dataset with variables that change more than once with a series of entry and exit dates



## The steps

- ▶ Split the first dataset (with just entry and exit sequentially by variables that change once or several times)
- ▶ Split by the time scale for LTMLE, e.g. 180 days
- ▶ Summarize for each period covariate status and outcome
- ▶ Move all covariates (not outcomes censoring and competing risk) one period back
- ▶ Transpose

