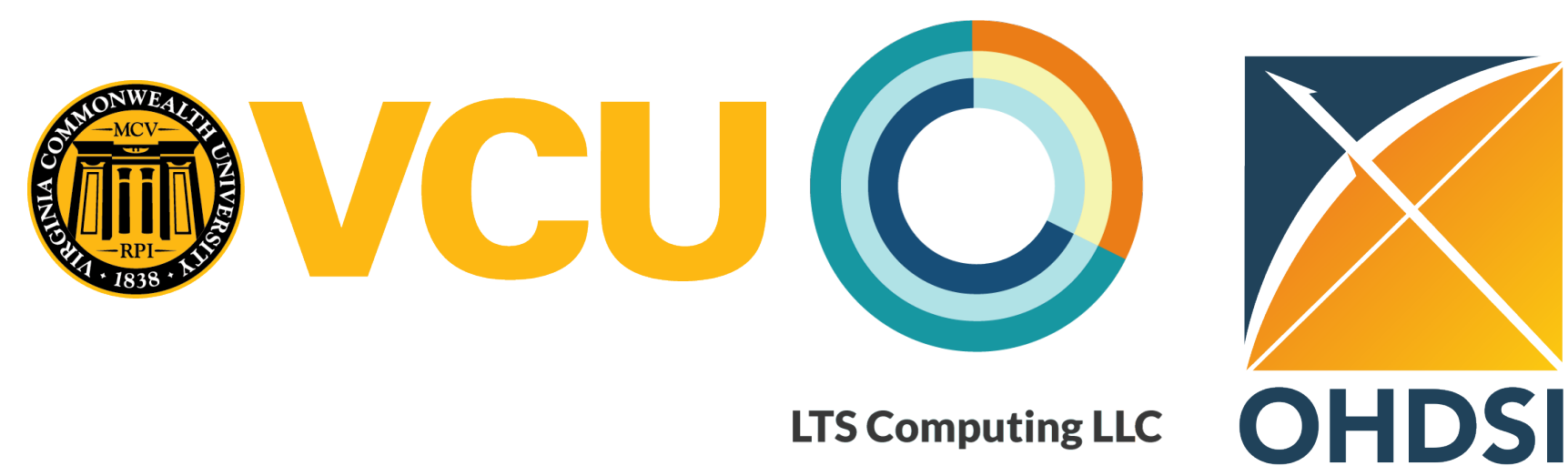# Introducing OHDSI Analytics at VCU in the LTS OHDSI cloud: An Example Application of Empirical Calibration Using Synthetic EMR Data for Detecting Cancer Risk in Oral Bisphosphonate Users

Martin Lavallee[1,2], Roy T. Sabo[1], Lee Evans[2] [1] Virginia Commonwealth University; Department of Biostatistics [2] LTS Computing LLC

INTRO:

- Using the OMOP CDM and OHDSI analytical tools provide a rigorous framework to generate evidence from EMR data
- Synthetic patient data converted to the OMOP CDM can be used as a training platform of methods without patient privacy concerns
- A goal was to evaluate synthetic EMR generator Synthea to see how well it performs as a space for testing and developing methods in the OHDSI framework
- Conducted example analysis on synthetic EMR dataset of Richmond VA and designed a training platform
- Introduce OHDSI using the training platform to VCU Biostatistics Department

METHODS:

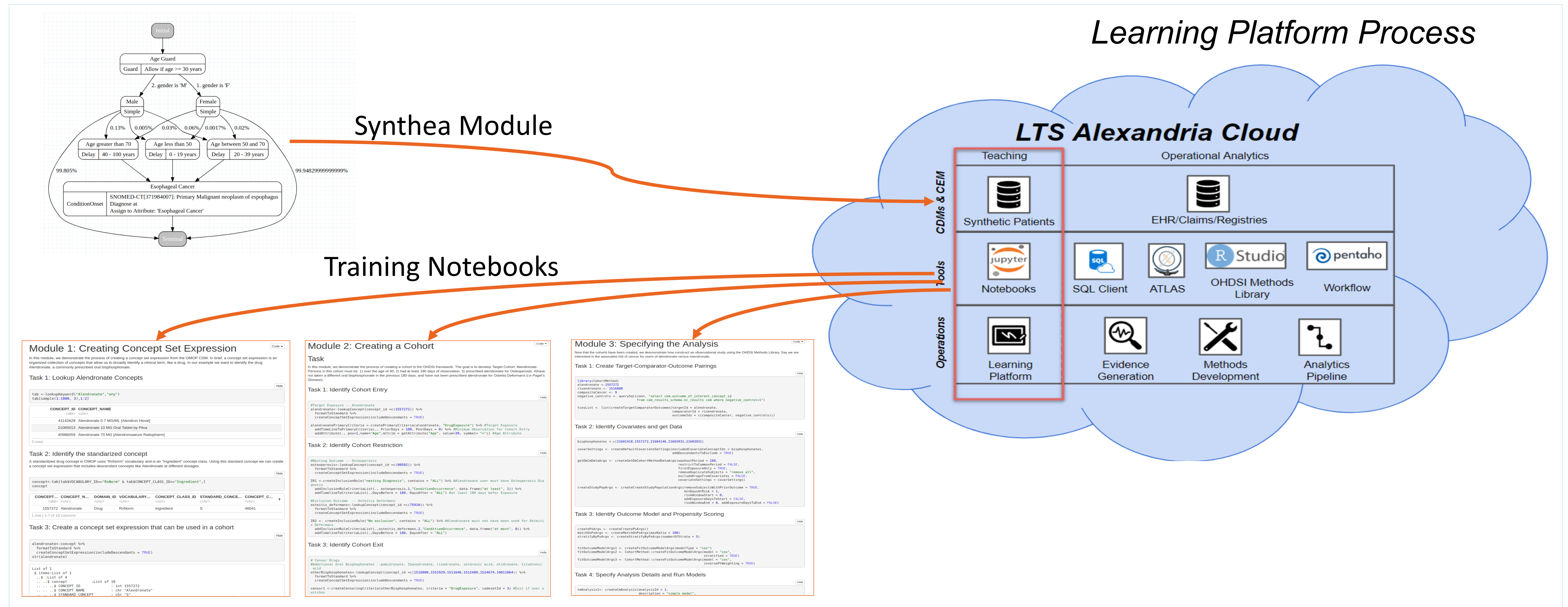### Learning Platform Process

- Create modules of cancer risk for esophageal and breast cancer and generate 1.3 million patients in Richmond VA
- Run ETL to convert to OMOP
- Load CEM data to query negative controls
- Generate ATLAS compatible cohorts using CAPR
- Set up Case-Control and SCCS analyses in R
- Design a series of teaching module Notebooks to demonstrate end-to-end process of conducting on OHDSI study
- Compare Synthea results to 5% sample of SYNPUF (n=116K)

Study Design

Design	Cohorts	Analytical Choices
Case-Control Emulates Green et al 2010	**T**: Alendronate **O**: Primary Malignant Neoplasm (colon, rectum, esophagus, stomach, breast) **Nesting**: Osteoporosis	- 5 cases per control - Match gender and age (caliper: 2) - Washout: 180 days - 68 drug negative controls from CEM
Self-Controlled Case Series	**T**: Alendronate **O**: Primary Malignant Neoplasm (colon, rectum, esophagus, stomach, breast) **Nesting**: Osteoporosis	- Naïve Period: 180 days - 68 drug negative controls from CEM

RESULTS:

- Using synthetic data for training in the OHDSI framework, we found our example study to show no association between oral bisphosphonates and composite cancer outcome
- Empirical calibration shows both methods are prone to systematic bias of the database. Synthea negative controls were harder to find
- SYNPUF 5% was a better synthetic dataset for answering the example question

Learning Platform Process

Synthea Module

Training Notebooks

LTS Alexandria Cloud

Module 1: Creating Concept Set Expression
Module 2: Creating a Cohort
Module 3: Specifying the Analysis

Lessons

- Goal was to prepare the learning platform process and create some example modules ready for introduction of the platform to VCU before the end of the year
- This clinical example was challenging because cancer is a difficult concept to define and may be rare in a study population
- Synthea requires development through addition of clinical modules that reflect real world practice and maintenance by validating and improving existing modules

Next Steps

- Work with clinical experts to improve and validate Synthea modules
- Continue to improve the Synthea ETL process
- On-going process of Introducing OHDSI to VCU Biostatistics Department
  - Overview of Learning Platform in October
  - Discussion of Empirical Calibration using negative controls in November
  - VCU to provide feedback on learning platform

Synpuf5pct

Design	RR	Conf. Int	Log(RR)	Se log(RR)
CC	1.094	(0.664, 1.718)	0.0899	0.243
SCCS	0.963	(0.586, 1.507)	-0.0378	0.241

SyntheaVA1M

Design	RR	Conf. Int	Log(RR)	Se log(RR)
CC	1.25	(0.358, 3.41)	0.223	0.575
SCCS	0.9433	(0.289, 2.24)	-0.058	0.522

Case-Control

Self-Controlled Case Series