Dynamic Evaluation of Cardiometabolic and Obesity DiseasE (DECODE) Model

Eric Wu, PhD[1], Jimmy Royer, PhD[2], Max Leroux, MSc[2], Intekhab Hossain, PhD[1], Robert W. Platt, PhD[3]

[1]Analysis Group, Inc, Boston, MA, USA; [2]Analysis Group, Inc, Montréal, QC, Canada; [3]McGill University, Montréal, QC, Canada

OBJECTIVES

- Develop a disease model to simulate the progression of obesity and related cardiometabolic conditions simultaneously
- Assess the long-term comprehensive clinical benefits of weight reduction

CONCLUSIONS

The DECODE model has demonstrated great potential to study various cardiometabolic conditions, interventions, outcomes and subpopulations flexibly.

The model estimated substantial benefits from weight loss in the obese population.

The model has successfully learned from different data sources and does not rely on (sometimes strong) assumptions of traditional models.

As a next step, the DECODE model is being trained on additional data sources and further validated with other long-term population cohort studies.

Poster presented at The Professional Society for Health Economics and Outcomes Research (ISPOR) Conference 13-16 May, 2025

ANALYSIS GROUP

BACKGROUND

Obesity and Cardiometabolic Conditions

- Obesity is a chronic, multifactorial condition associated with increased risk of over 200 comorbidities, leading to substantial healthcare utilization and costs.[1,2]
- Cardiometabolic conditions like obesity and T2DM are interconnected disorders that share common biological pathways and risk factors, with complex interactions as they progress.
- Quantifying the long-term comprehensive downstream impacts of anti-obesity interventions such as GLP-1 remains a challenge due to
  - The complex interaction among the cluster of conditions
  - The lack of long-term comprehensive data
  - Complexity in assessing combined benefits of GLP-1s (e.g. weight reduction and A1C reduction)
  - Healthcare decision makers often need to evaluate benefits across a wide range of patient subpopulations

Methodology

- Advancements in generative AI (GenAI), such as in simulating patient digital twins, are unlocking powerful new capabilities to understand complex diseases holistically with great depth and precision. GenAI models:
  - can incorporate multiple data sources to develop a comprehensive understandings of diseases
  - do not rely on oversimplified assumptions like many traditional models do
  - are flexible to studying a wide range of research questions over diverse subpopulations and studying the benefits of treatments with multiple related indications
- Regulatory bodies have been receptive to the exploration of AI-simulated data use in HEOR applications
  - In Sept. 2022, the **EMA** permitted digital twin models for use in Phase II and III clinical trials.[3]
  - In Mar. 2024, the **FDA**, **NIH**, and **NSF** created the *Foundations for Digital Twins as Catalyzers of Biomedical Technological Innovation* (FDT-BioTech) program to advance biomedical innovation through the development of algorithms relevant to digital twins and synthetic humans.[4]

METHODS

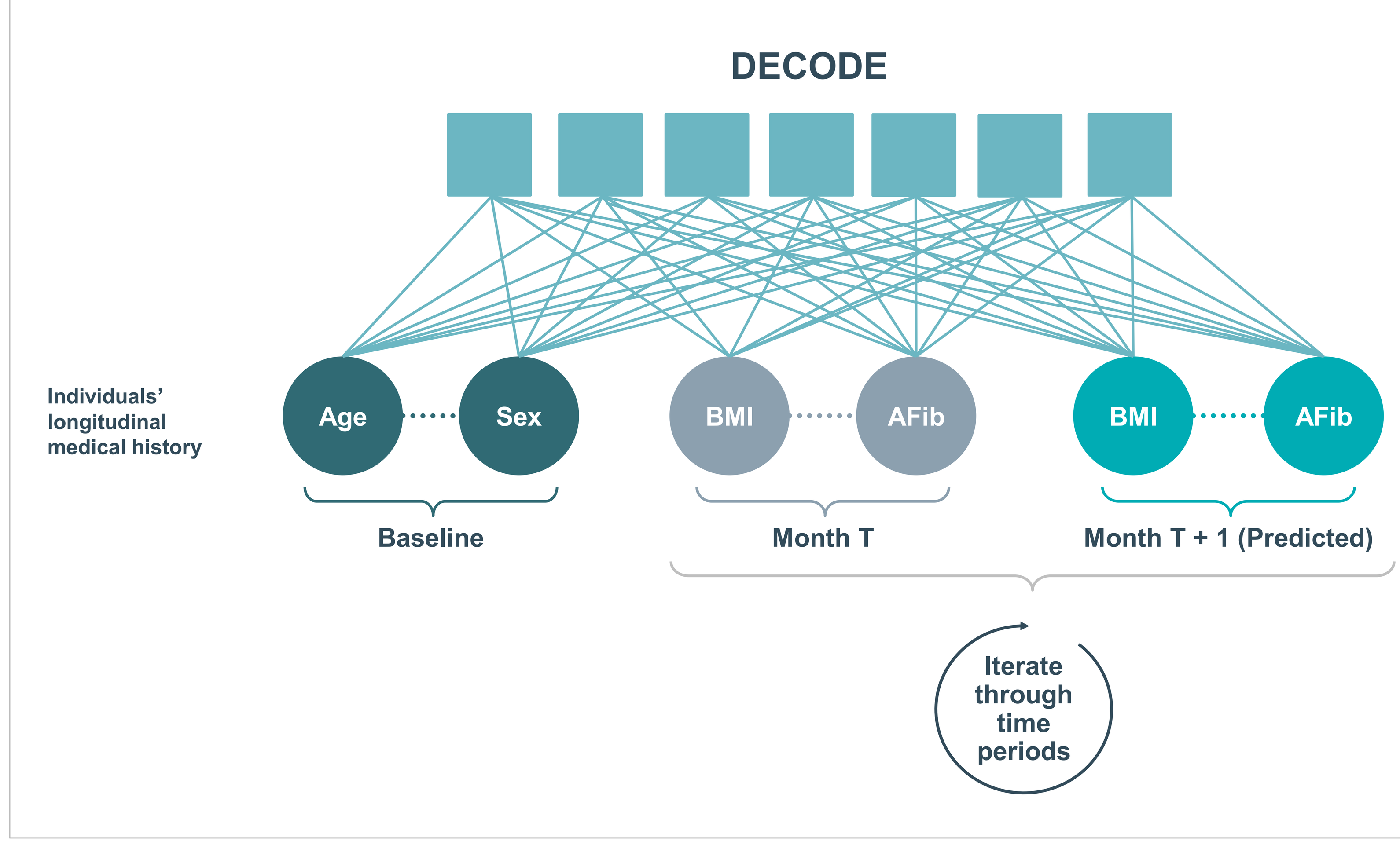
- Conditional Restricted Boltzmann Machines (CRBMs) are deep learning models that uncover hidden relationships in data by modeling patterns between observed inputs and latent features.[5]
- Two large US databases – one EHR and one claims – were jointly used to train a digital twin Dynamic Evaluation of Cardiometabolic and Obesity DiseasE (DECODE) model to understand long-term progression of obesity, cardiometabolic, and other related conditions (such as cardiovascular conditions, T2DM, hypertension, dyslipidemia, MASH, obstructive sleep apnea, and knee osteoarthritis)
- Adult lives over the age of 18 with sufficient longitudinal clinical data on BMI, lab values, and cardiometabolic conditions were selected.
- The model considered both fixed baseline variables and time-varying variables.

The DECODE model was trained on the following two data sources, resulting in 18,636 parameters

- **Dandelion Health's EHR data**
  - The database includes rich long-term clinical details of over 10 million lives from three major US health systems
  - The data comprises of variables within six clinical categories: medical diagnoses, procedures, medications, labs, vitals, and imaging.
- **Claims data**
  - A large commercial database of US medical claims of over 170 million lives.
- Model performance was assessed across a 5-year time horizon on a test cohort of 10,000+ lives from the Dandelion dataset.

Dandelion	US Claims
Any of diabetes, obesity/overweight, ASCVD, MASH, HFpEF, Alzheimer's disease, COPD, asthma, PH/PAH, CKD **+20% lives with none of the above**	Any of diabetes, obesity/overweight, GLP-1 NDC
June 2006 – Mar 2024	**Jan 2016 – Sep 2024**
Lives with at least 1 year of continuous activity & sufficient longitudinal data across all modeled variables	Random representative sample of claims & sufficient longitudinal data across all modeled variables
Final cohort of 228,756 lives	**Final cohort of 240,155 lives**
Variables 53 time-dependent 9 static baseline **Lives follow-up periods** ranged 1-18 years **Time period** 2007 – 2024	**Variables** 48 time-dependent 5 static baseline **Lives follow-up periods** ranged 1-8 years **Time period** 2016 – 2023

Figure 1. Visualization of the model



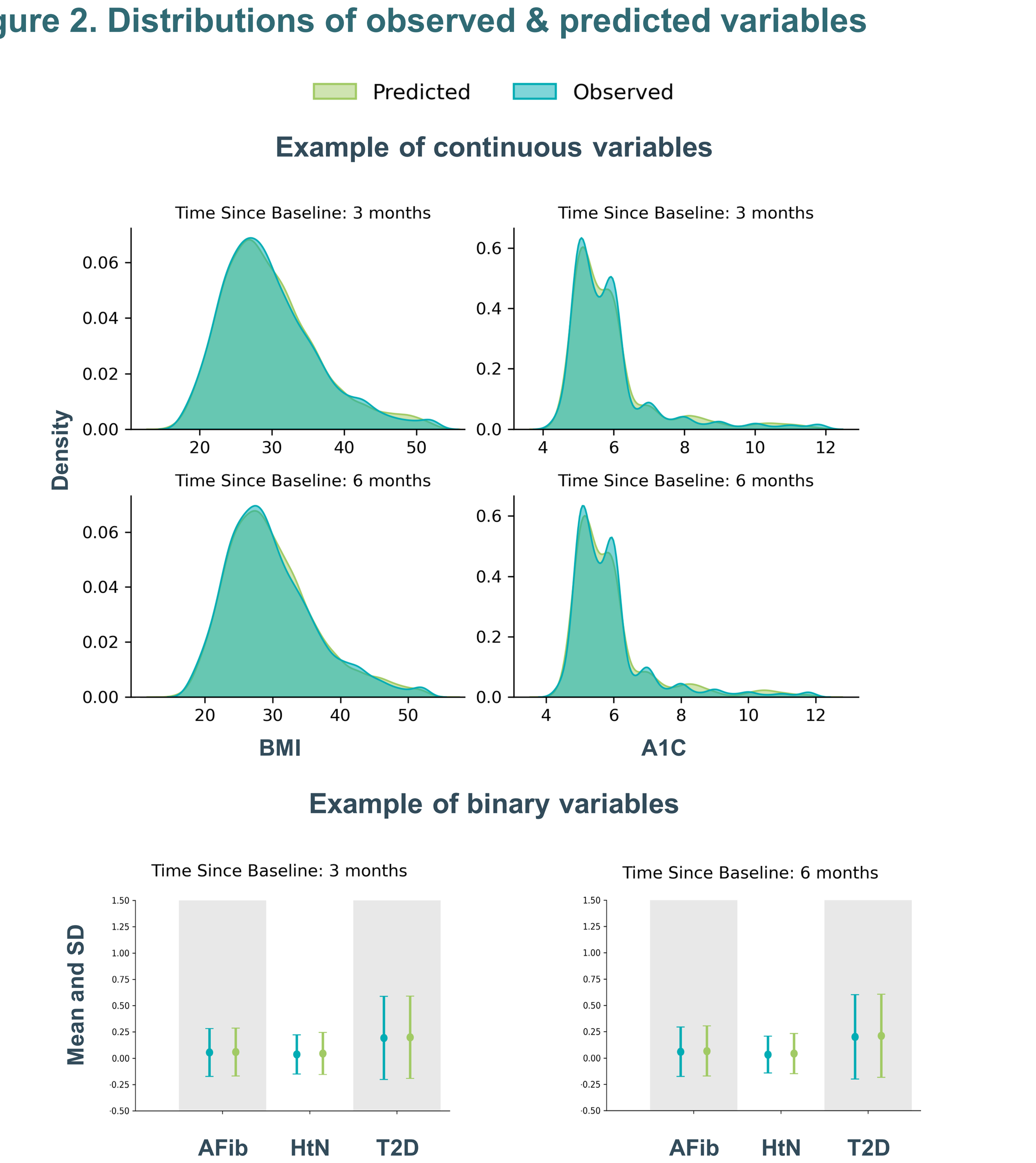
RESULTS

Model validation

Method 1: Distributions of each variable at each time point were compared between the observed & predicted data

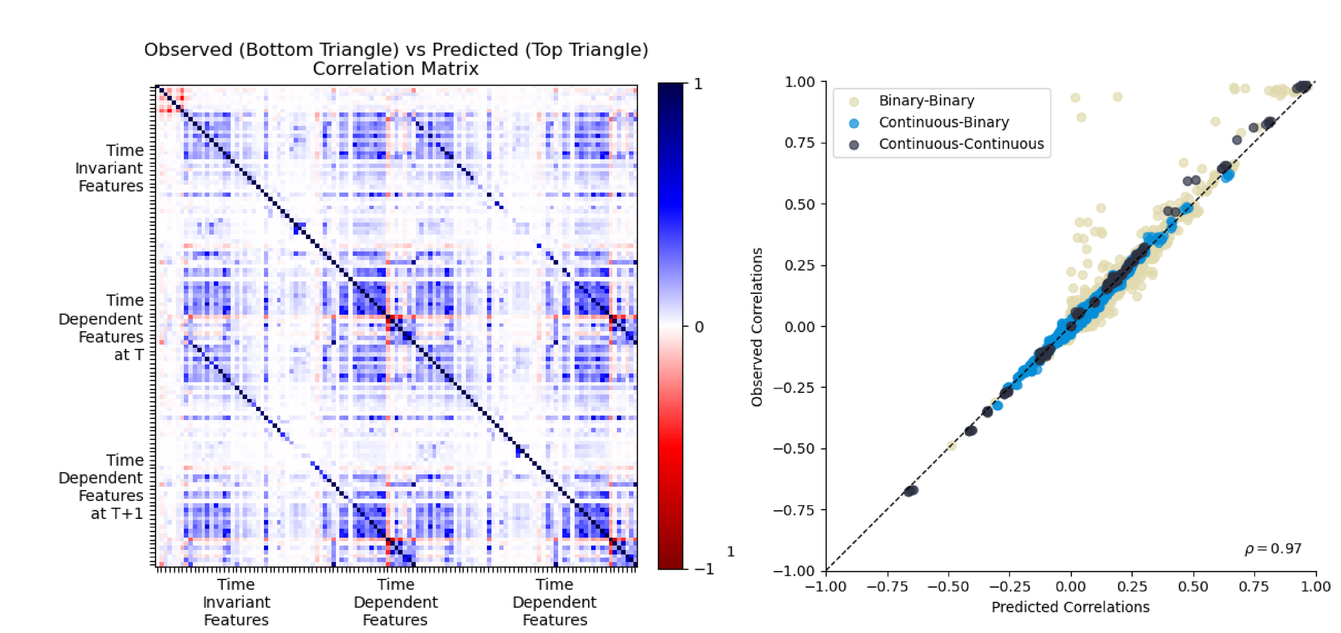
- The model showed strong correlation between observed and predicted distributions, in terms of binary proportions (ρ=0.99), continuous means (ρ=0.99), variances (ρ=0.99) and covariance structure (ρ=0.97).
- The mean Intersection over Union (IoU)[6] - a non-parametric measure of the closeness between two distributions ranging from **0 (no overlap)** to **1 (perfect match)** - across all time points between observed and predicted distributions was **0.89 for binary variables and 0.90 for continuous variables.**

Figure 2. Distributions of observed & predicted variables



Method 2: Correlations between variables, including across timepoints were compared between the observed & predicted data[7]

Figure 3. Correlation between observed and predicted correlations ρ=0.97



Model validation on subgroups

Subpopulation	Number of test set patients	Method #1: IoU binary variables	Method #1: IoU cont. variables	Method #2: Correlation of correlations
Baseline BMI >= 40	985	0.86	0.81	0.91
Baseline Age >= 65	2763	0.83	0.88	0.94
T2DM at baseline	2310	0.82	0.87	0.89
GLP1 prescribed at any point	2085	0.80	0.87	0.93

Estimating long-term benefits of weight loss

- The DECODE model was used to predict benefits of a 10% weight loss in an obese population

In year **ten,** the risk ratios (RRs) are:

Heart failure	Atrial fibrillation	Bariatric Surgery
0.79	0.84	0.61

In year **five,** the RRs are:

Heart failure	Atrial fibrillation	Bariatric Surgery
0.81	0.84	0.47

REFERENCES

1. Powell-Wiley, Tiffany M et al. "Obesity and Cardiovascular Disease: A Scientific Statement From the American Heart Association." *Circulation* vol. 143,21 (2021): e984-e1010.
2. le Roux, Carel W et al. "Obesity, cardiovascular risk and healthcare resource utilization in the UK." *European journal of preventive cardiology* vol. 28,11 (2021): 1235-1241.
3. European Medicines Agency, "Qualification opinion for Prognostic Covariate Adjustment (PROCOVA™)", *Qualification* (2022), EMADOC-1700519818-907465.
4. U.S. National Science Foundation, "NSF 24-561: Foundations for Digital Twins as Catalyzers of Biomedical Technological Innovation", *Solicitation* (2024).
5. Fisher, Charles K., Aaron M. Smith, et al. "Machine learning for comprehensive forecasting of Alzheimer's Disease progression" Scientific reports 9, no. 1 (2019): 13622.
6. Everingham, M., Van Gool, L., Williams, C.K.I. et al. "The Pascal Visual Object Classes (VOC) Challenge" Int J Comput Vis 88, (2010): 303–338.
7. Ramsay, James et al. "Matrix Correlation", *Psychometrika* 49.3 (1984): 403-423.