

Mathematical Modeling in Precision Nutrition

Diana Thomas, PhD



Artificial Intelligence, Data Engineering and Machine Learning (AIDE-ML) Center



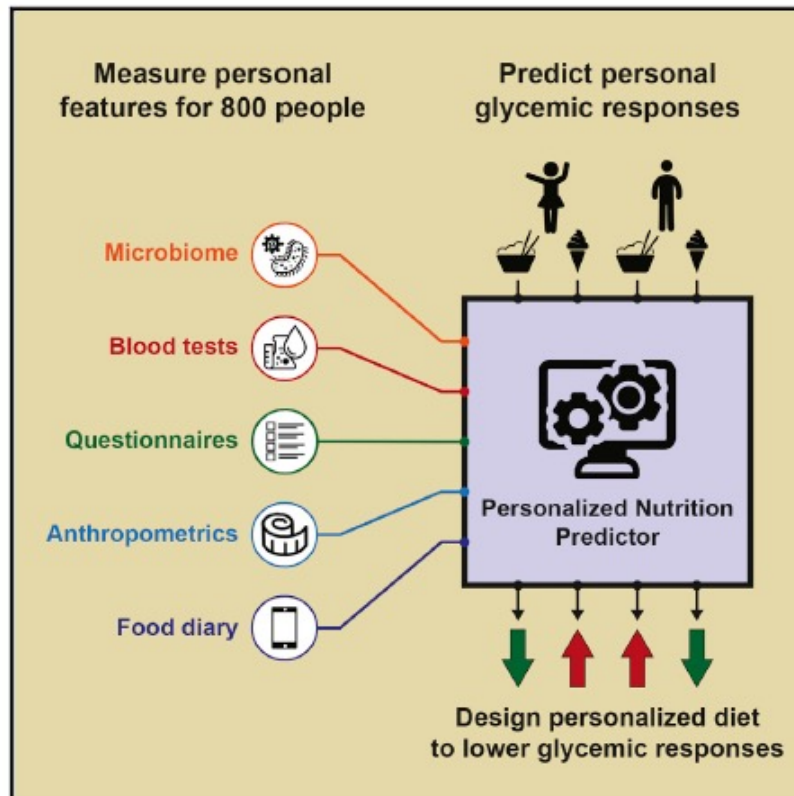
Thank you to our funding
NIH IAA AOD22022001



Cell

Personalized Nutrition by Prediction of Glycemic Responses

Graphical Abstract



Authors

David Zeevi, Tal Korem, Niv Zmora, ...,
Zamir Halpem, Eran Elinav, Eran Segal

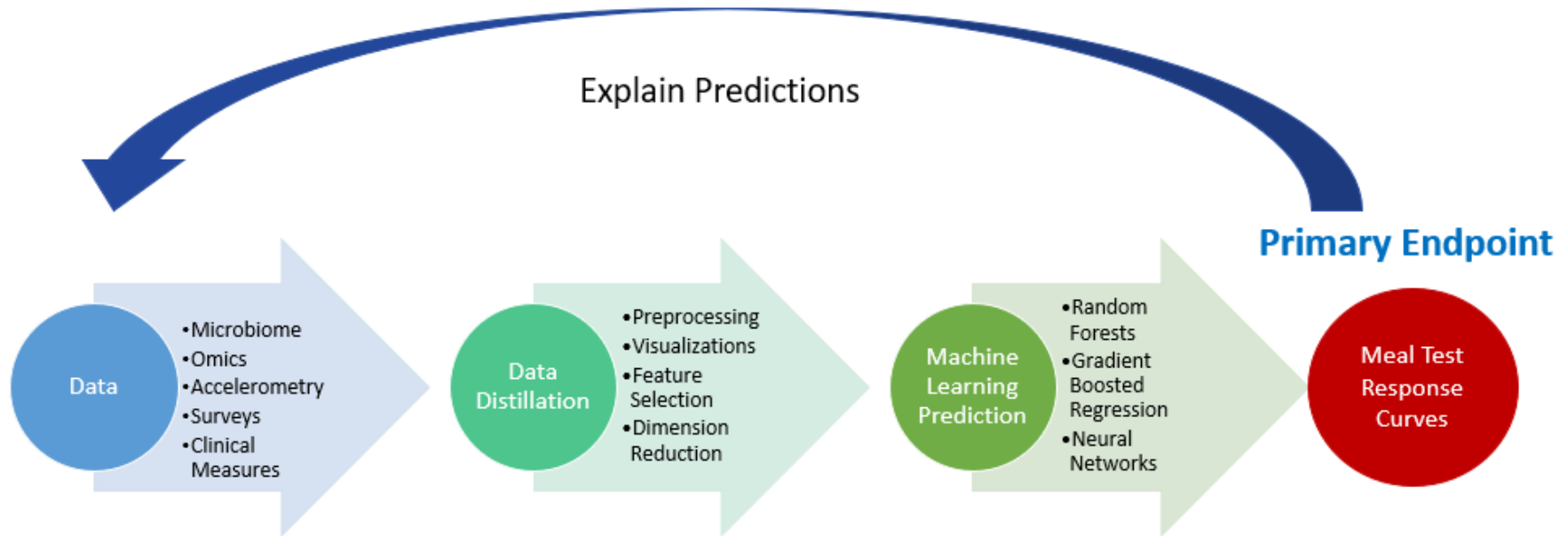
Correspondence

eran.elinav@weizmann.ac.il (E.E.),
eran.segal@weizmann.ac.il (E.S.)

In Brief

People eating identical meals present high variability in post-meal blood glucose response. Personalized diets created with the help of an accurate predictor of blood glucose response that integrates parameters such as dietary habits, physical activity, and gut microbiota may successfully lower post-meal blood glucose and its long-term metabolic consequences.

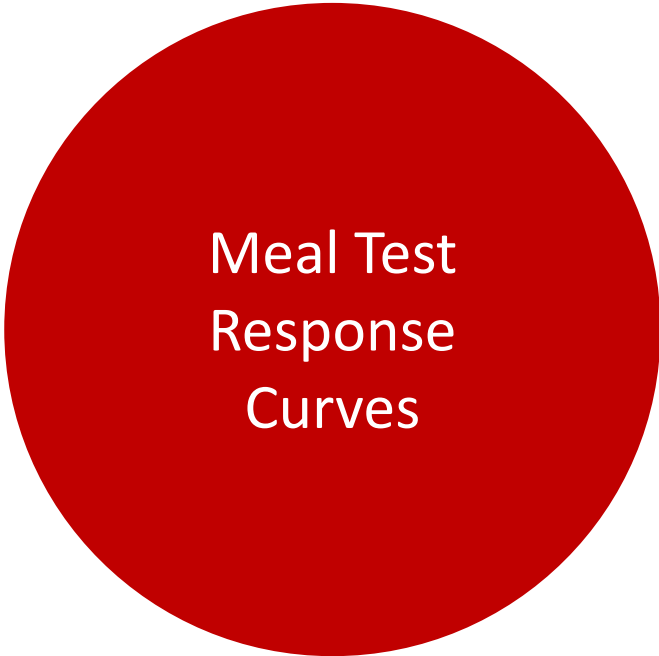
How does the Modeling Work?



How does the Modeling Work?

The Primary Endpoint

Primary Endpoint

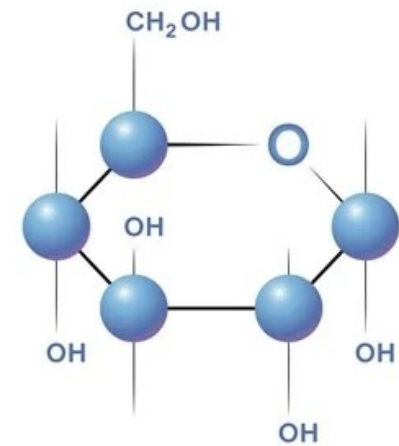


Meal Test
Response
Curves

Motivation for Primary Endpoint

- Post-prandial glucose response (PPGR) curves were the primary endpoint.
- There is an expected post-prandial response to a meal which will vary by individual.
- Post-prandial hyperglycemia may be a risk factor for a whole host of diseases.

This was a **discovery science** project that searched for factors that may explain interpersonal differences in response (PPGR) coming from an array of data.



Glucose



Walter Bradford Cannon



Born October 19, 1871

Homeostasis is the state of conditions maintained by humans.

Examples include body temperature, energy balance, and blood sugar level.

The theory of homeostasis suggests that individuals with **post-prandial response to meals that deviate more from homeostasis than “normal”** is a risk factor.

Motivation for Primary Endpoint



Comes from the oral glucose tolerance test (OGTT).

Higher curve displacement is associated with higher risk of diabetes and cardiovascular disease

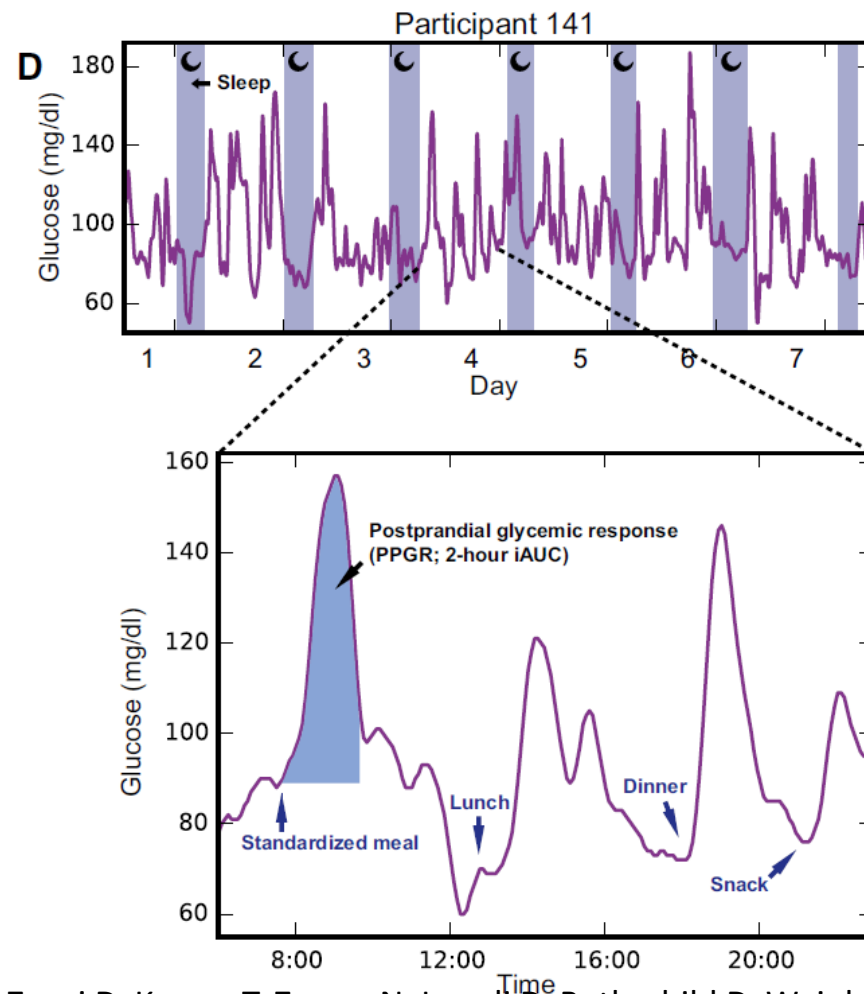
Additional post-prandial analytes such as triglycerides have been measured and demonstrated predictive of metabolic risk

The mixed meal tolerance test (MMTT) extends the concept of the OGTT: liquid mixed macronutrient drink or solid mixed macronutrient foods (e.g. bread, rice, mixed meal, muffins)



How to measure “deviate more than normal”

- iPro2 CGM worn for 7 days
- Frequency is every 5 minutes
- Measurements retained over a 2 hour period after standardized meal
- iAUC was the collapse of the PPGR curve



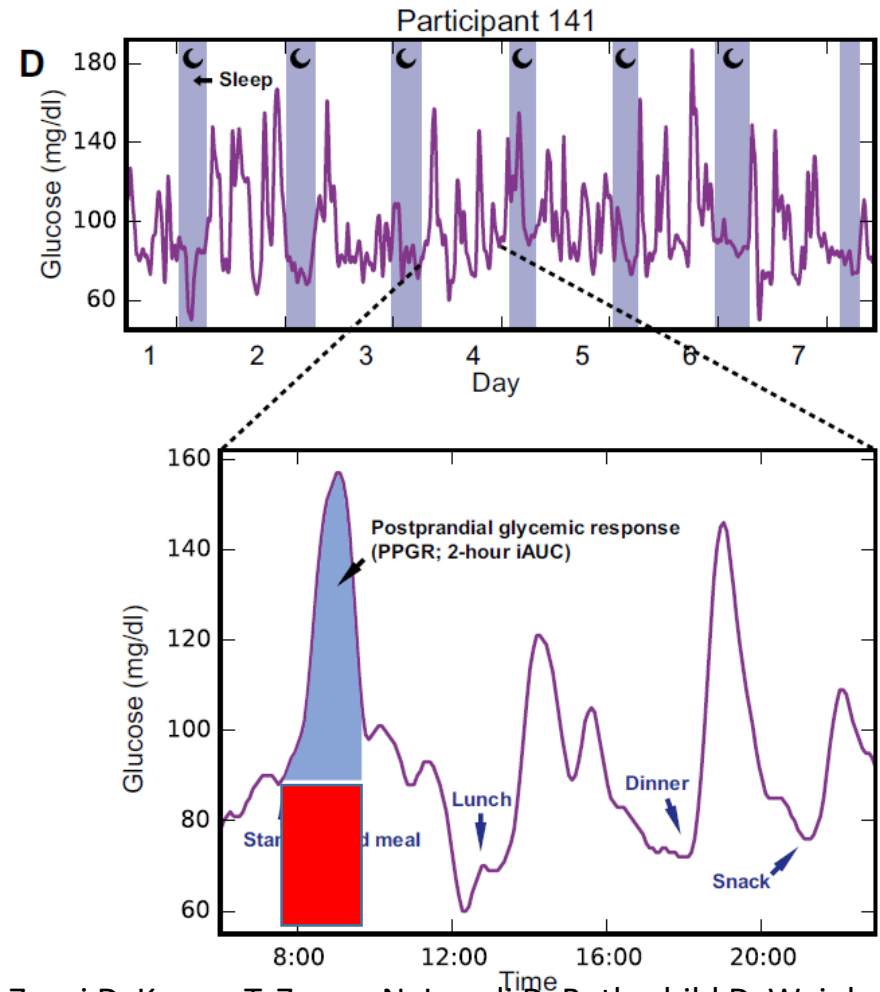
Personalized Nutrition by Prediction of Glycemic Responses. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, Ben-Yacov O, Lador D, Avnit-Sagi T, Lotan-Pompan M, Suez J, Mahdi JA, Matot E, Malka G, Kosower N, Rein M, Zilberman-Schapira G, Dohnalová L, Pevsner-Fischer M, Bikovsky R, Halpern Z, Elinav E, Segal E. Cell. 2015 Nov 19;163(5):1079-1094. doi: 10.1016/j.cell.2015.11.001.



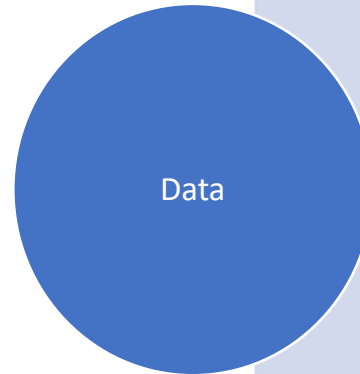
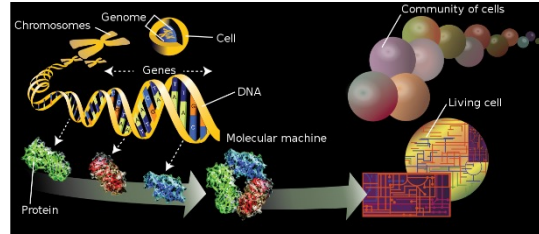
How to measure “deviate more than normal”

iAUC is the difference between the total area under the curve and the red rectangle.

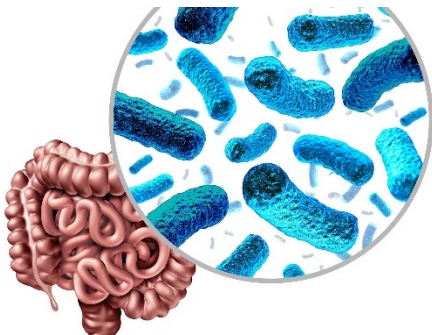
Blue area is the deviation from homeostasis.



Personalized Nutrition by Prediction of Glycemic Responses. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, Ben-Yacov O, Lador D, Avnit-Sagi T, Lotan-Pompan M, Suez J, Mahdi JA, Matot E, Malka G, Kosover N, Rein M, Zilberman-Schapira G, Dohnalová L, Pevsner-Fischer M, Bikovsky R, Halpern Z, Elinav E, Segal E. *Cell*. 2015 Nov 19;163(5):1079-1094. doi: 10.1016/j.cell.2015.11.001.



- Microbiome
- Omics
- Accelerometry
- Surveys
- Clinical Measures





Data Distillation

- Preprocessing
- Visualizations
- Feature Selection
- Dimension Reduction

Making the data
“AI ready”



Data Preprocessing

Data Exploration: Understanding Data/Quality Assurance

Feature Selection: Clustering/PCA/Recursive

Iterative “Sniff Checks”: Ensure that your data is free from errors and ethical issues

DATA-DRIVEN TO MADNESS

Making the data “AI ready”

Bigger is not always better!


- Including model inputs that are superfluous and do not add new information to the dataset can lead to errors in predictive models.
- We need to collapse and transform nutrition-relevant data to create datasets that are AI-ready and ready for a range of other analyses.

WHAT TO DO WHEN THERE'S

'TOO MUCH' DATA



Machine learning modeling practices to support the principles of AI and ethics in nutrition research

[Diana M. Thomas](#) , [Samantha Kleinberg](#), [Andrew W. Brown](#), [Mason Crow](#), [Nathaniel D. Bastian](#), [Nicholas Reisweber](#), [Robert Lasater](#), [Thomas Kendall](#), [Patrick Shafto](#), [Raymond Blaine](#), [Sarah Smith](#), [Daniel Ruiz](#), [Christopher Morrell](#) & [Nicholas Clark](#)

[Nutrition & Diabetes](#) **12**, Article number: 48 (2022) | [Cite this article](#)

1806 Accesses | **2** Citations | **10** Altmetric | [Metrics](#)



NO “PUSH AND PLAY” FOR AI!

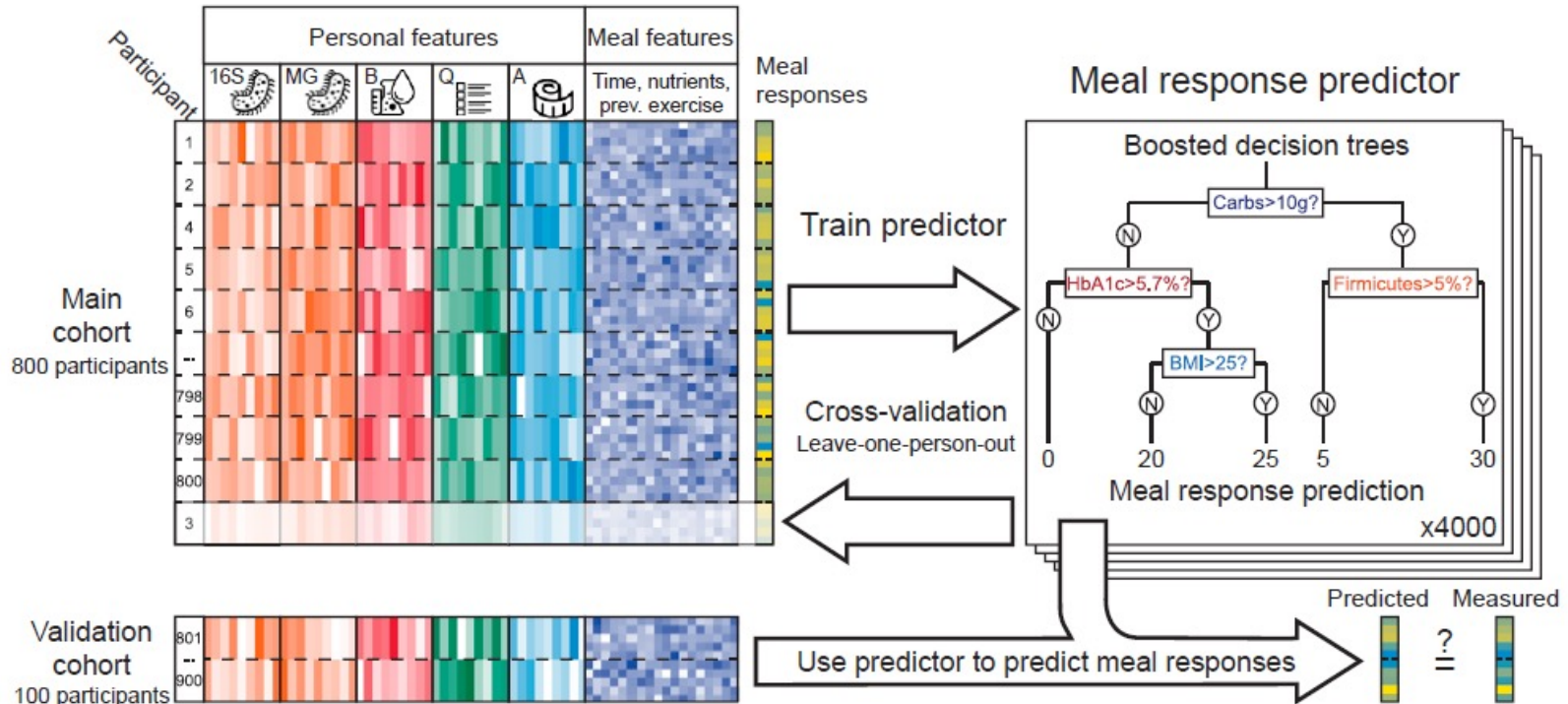


Machine
Learning
Prediction

- Random Forests
- Gradient Boosted Regression
- Neural Networks



The actual predictive algorithm



Trained on 800 with leave out one cross validation (R=0.68) -tested in 100 independent samples (R=0.70)-Note R=0.71-0.77 was the intra-individual association.

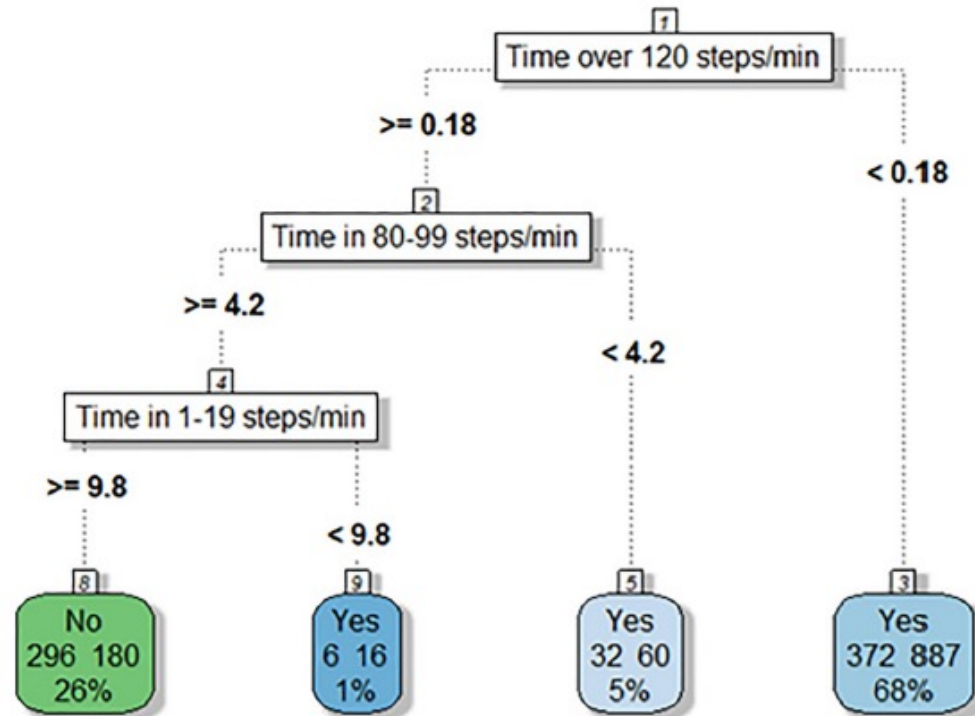
Zeevi D, et al. Personalized Nutrition by Prediction of Glycemic Responses. Cell. 2015 Nov 19;163(5):1079-1094.

Adams B, Fidler K, Demoes N, Aguiar EJ, Ducharme SW, McCullough AK, Moore CC, Tudor-Locke C, Thomas D. Cardiometabolic thresholds for peak 30-min cadence and steps/day. PLoS One. 2019

Why Decision Trees?

Decision trees represent ML models that predict in a way that you can visualize and clinically understand.

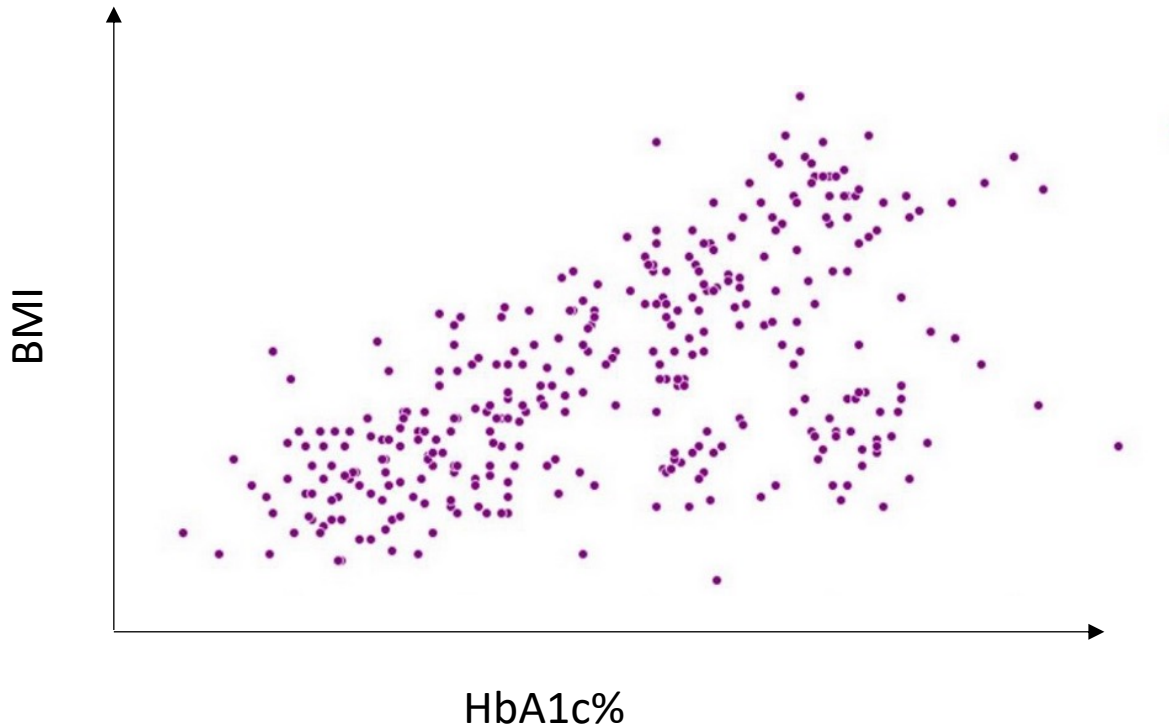
Decision tree prediction however is often weaker than other types of ML models like neural networks.



Zeevi D, et al. Personalized Nutrition by Prediction of Glycemic Responses. Cell. 2015 Nov 19;163(5):1079-1094.

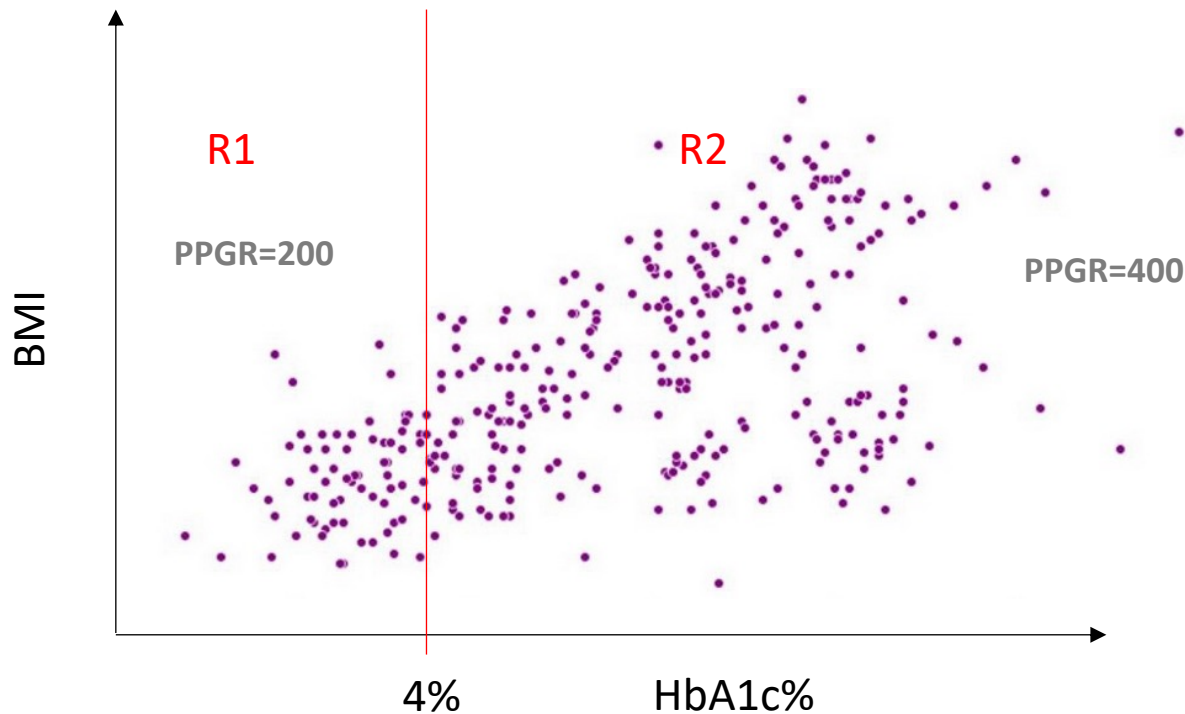


A made up toy example: Suppose your trying to predict PPGR iAUC from BMI and HbA1c



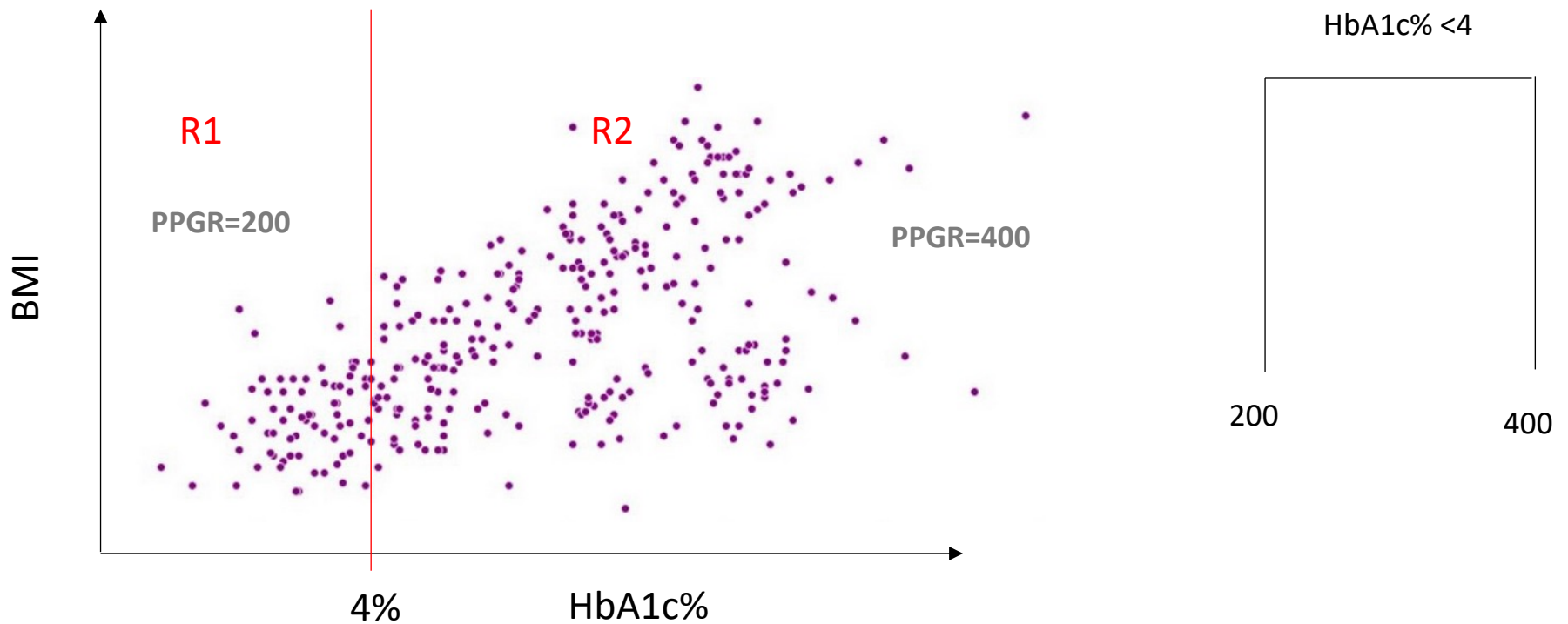


A made up toy example: Suppose your trying to predict PPGR iAUC from BMI and HbA1c



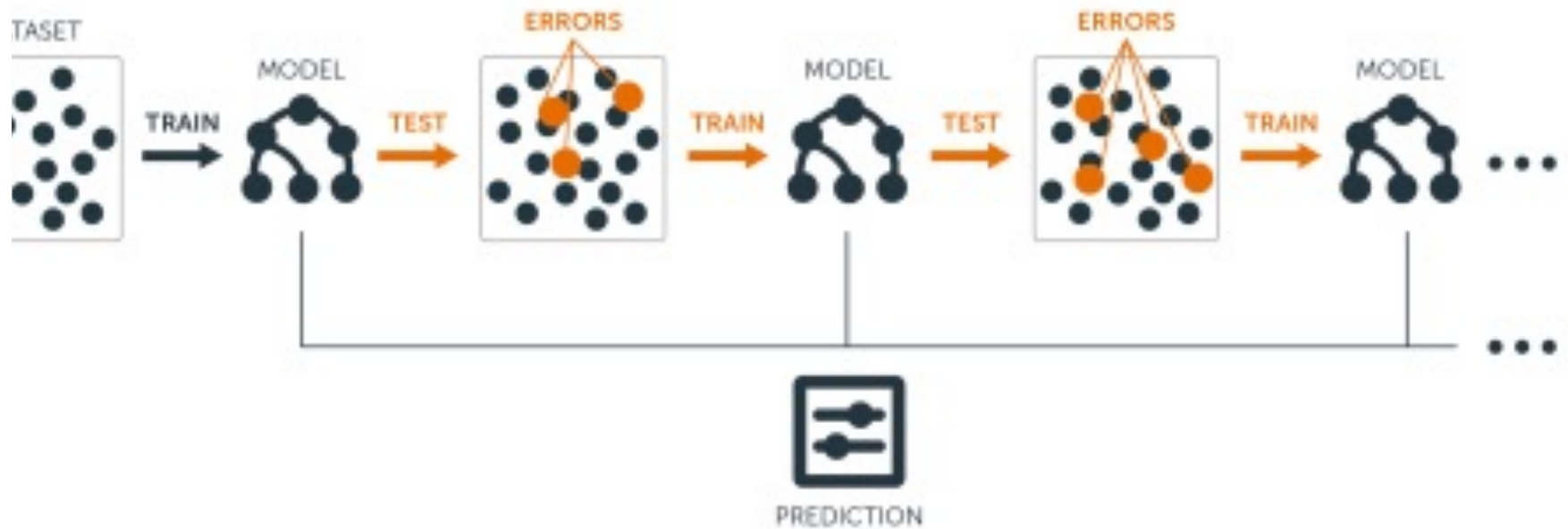


A made up toy example: Suppose your trying to predict PPGR iAUC from BMI and HbA1c



Build a regression tree to predict the residuals from the input variables.

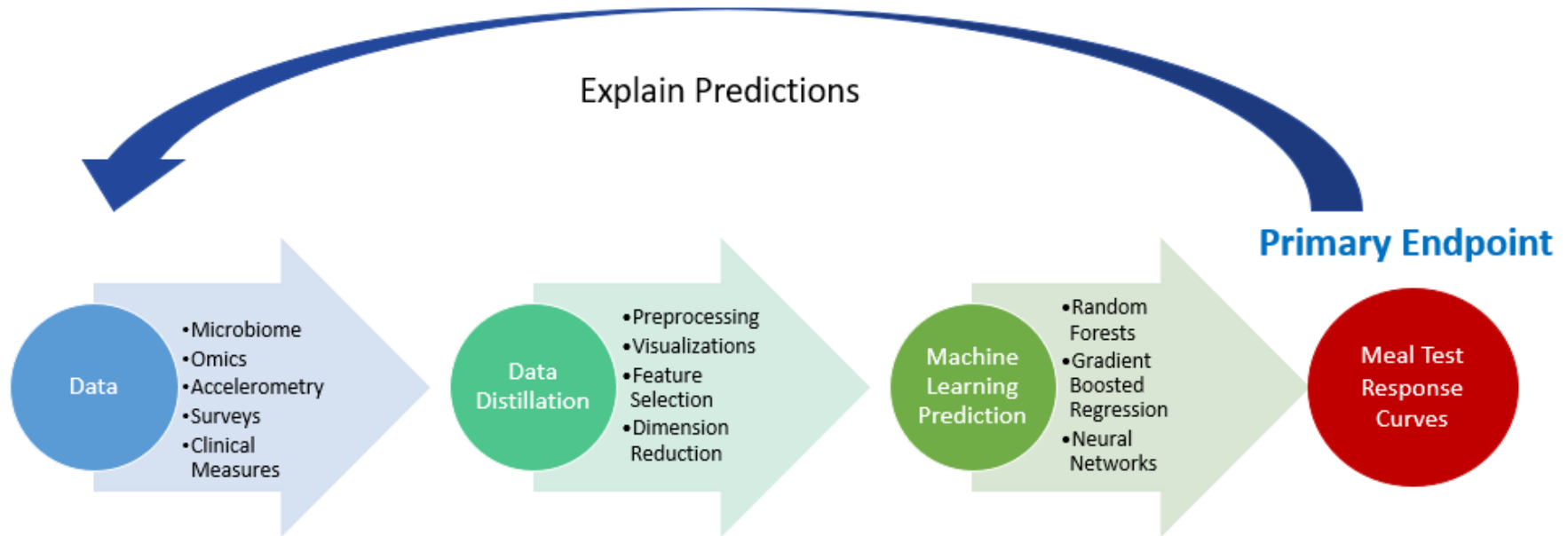
- Many decision trees that are “weak predictors”
- Iteratively build up strong predictors from these weak predictors.





used as independent variables to predict each internal molecule. The random forest model (R package 'caret' and 'RandomForest'), which has been proven to have the best prediction accuracy, was used³⁰. The

- Decision tree-based algorithms are preferred because you can input ordinal data (Likert Survey data) AND continuous data together.
- Some machine learning models are designed just for ordinal categorical data (PLS-DA) and some are designed to handle only continuous data (neural networks).
- Best practices: you need to try several models including the baseline one.





DEFENSE ADVANCED
RESEARCH PROJECTS AGENCY

There are two products from this exercise:

The Explainable Algorithm

- **The new previously unknown factors that impact the response to diets.**
- The algorithm itself.