Evaluating the Impact of Dietary Supplement Use on Health Outcomes Using Causal Machine Learning

Karthik Vakada
Department of Computer Science
University of Maryland
College Park, USA

Sri Akash Kadali
Department of Computer Science
University of Maryland
College Park, USA

***Abstract*—Estimating causal effects from observational health data is challenging due to confounding, selection bias, and heterogeneous treatment responses. This work studies the causal impact of dietary supplement use on cardiometabolic health outcomes using data from the National Health and Nutrition Examination Survey (NHANES) 2021–2023. We frame supplement use as a binary treatment and analyze its effects on mean systolic blood pressure and fasting plasma glucose. To address confounding, we apply a Double Robust Learner implemented via EconML, combining flexible machine learning models for outcome and propensity estimation with orthogonalized treatment effect estimation. We report average treatment effects with bootstrap confidence intervals and estimate individual-level conditional average treatment effects to study heterogeneity across demographic and baseline health subgroups. Results are compared against a classical ordinary least squares baseline with robust standard errors to highlight differences between associative and causal estimates. The entire analysis is implemented in a modular, reproducible pipeline with a clear separation between data preprocessing, causal estimation, and evaluation, and is fully reproducible via a Docker-based execution environment. This project demonstrates principled causal inference on real-world observational health data and emphasizes correct causal framing over predictive performance.**

***Index Terms*—Causal inference, EconML, NHANES, treatment effects, healthcare analytics**

I. Introduction

Understanding the effects of health interventions using real-world data is a central challenge in modern data science and applied machine learning. Large-scale observational datasets, such as national health surveys, offer rich information on patient characteristics, behaviors, and outcomes, but they do not arise from randomized experiments. As a result, naïve statistical analyses often conflate correlation with causation, leading to biased or misleading conclusions when estimating the impact of interventions.

Dietary supplement use is a representative example of this challenge. Individuals who take supplements differ systematically from those who do not along dimensions such as age, socioeconomic status, baseline health, and health-seeking behavior. These differences introduce confounding that invalidates simple comparisons of outcomes between treated and untreated groups. Consequently, estimating the causal effect of supplement use requires methods that explicitly account for non-random treatment assignment while remaining robust to high-dimensional covariates.

Recent advances in causal machine learning provide tools designed to address these issues by combining flexible predictive models with principled causal estimation. In particular, double robust estimators allow consistent estimation of treatment effects when either the outcome model or the treatment assignment model is correctly specified. The EconML library operationalizes these ideas by providing reusable implementations of double machine learning estimators that are suitable for applied settings.

In this project, we apply causal machine learning methods to estimate the effect of dietary supplement use on selected cardiometabolic health outcomes using data from the National Health and Nutrition Examination Survey (NHANES) 2021–2023. We focus on mean systolic blood pressure and fasting plasma glucose as outcomes of interest and treat supplement use as a binary intervention. Our analysis emphasizes causal estimation rather than prediction and reports both average treatment effects and individual-level conditional effects to explore treatment effect heterogeneity.

Beyond estimation, this work places strong emphasis on reproducibility and methodological clarity. The analysis pipeline is structured into distinct layers for data preprocessing, causal estimation, and evaluation, and is fully executable in a containerized environment. By combining transparent baselines with modern causal machine learning techniques, this project illustrates how principled causal inference can be performed on real-world observational health data while maintaining reproducibility and interpretability.

II. Problem Statement and Objectives

The central problem addressed in this work is the estimation of causal effects of a health-related intervention using observational data. In the absence of randomized assignment, individuals who receive an intervention may differ systematically from those who do not, leading to confounding that invalidates naïve statistical comparisons. This challenge is particularly acute in population-level health surveys, where treatment decisions

are driven by personal behavior, access to healthcare, and underlying health status rather than experimental design.

Within this context, dietary supplement use serves as a representative intervention of interest. Supplement consumption is widespread and self-selected, and individuals who report supplement use often differ from non-users in age, baseline health indicators, and health-seeking behavior. As a result, estimating the effect of supplement use on health outcomes requires methods that can adjust for observed confounders while remaining robust to model misspecification and high-dimensional covariates.

The primary objective of this project is to estimate the causal impact of dietary supplement use on selected cardiometabolic outcomes using data from the National Health and Nutrition Examination Survey (NHANES) 2021–2023. Specifically, we aim to quantify the average effect of supplement use on mean systolic blood pressure and fasting plasma glucose while accounting for observed demographic and clinical covariates.

A secondary objective is to examine heterogeneity in treatment effects across individuals and subpopulations. Rather than relying solely on average treatment effects, this work seeks to characterize variation in causal effects as a function of baseline characteristics, providing insight into whether the estimated impact of supplement use differs across demographic or health-related subgroups.

Finally, this project aims to demonstrate a reproducible and modular causal inference workflow suitable for applied machine learning settings. By structuring the analysis into clear data preparation, causal estimation, and evaluation layers, and by comparing causal machine learning estimates with a classical regression baseline, the project emphasizes methodological rigor, transparency, and reproducibility over predictive performance.

III. Dataset Description

This study uses data from the National Health and Nutrition Examination Survey (NHANES) for the 2021–2023 survey cycle. NHANES is a nationally representative, cross-sectional survey conducted by the Centers for Disease Control and Prevention that combines interviews, physical examinations, and laboratory measurements to assess the health and nutritional status of the U.S. population. The survey employs a complex, multistage probability sampling design to ensure population-level coverage across demographic groups.

The analysis integrates multiple NHANES component datasets at the individual respondent level using the unique participant identifier. Demographic information is obtained from the demographics component, while clinical and laboratory measurements are sourced from examination and laboratory components. Dietary supplement use is derived from questionnaire-based supplement intake records. All datasets are merged to form a unified analysis-ready table, with one row per respondent.

Key variables used in this project include demographic characteristics such as age and sex, anthropometric measures including body mass index and body weight, cardiometabolic biomarkers such as systolic blood pressure and fasting plasma glucose, lipid profile measures, and inflammatory markers. Dietary supplement use is encoded as a binary indicator reflecting whether a respondent reports using any dietary supplement during the survey period.

Prior to analysis, datasets undergo preprocessing to ensure consistency and validity. This includes harmonizing variable names across components, converting measurements to numeric form where appropriate, handling missing values, and removing implausible observations based on clinically reasonable ranges. Feature construction and validation are centralized in a reusable utility module to ensure consistency across analyses.

Because NHANES is observational and cross-sectional, treatment assignment is not randomized, and outcomes and covariates are measured contemporaneously. As a result, causal interpretation relies on adjusting for observed confounders rather than experimental control. This dataset structure motivates the use of causal machine learning methods capable of addressing confounding while leveraging the richness of the available covariates.

A. NHANES Overview

The National Health and Nutrition Examination Survey (NHANES) is a continuous, cross-sectional survey conducted by the Centers for Disease Control and Prevention to assess the health and nutritional status of the civilian, non-institutionalized population of the United States. NHANES combines self-reported questionnaire data with standardized physical examinations and laboratory measurements collected through mobile examination centers, providing a comprehensive view of population health.

The survey follows a complex, multistage probability sampling design that oversamples specific demographic groups to improve estimation accuracy for underrepresented populations. As a result, NHANES data are commonly used for population-level health analysis, epidemiological studies, and policy-relevant research. However, the data are observational in nature, and individuals are not randomly assigned to behaviors or interventions such as dietary supplement use.

NHANES data are organized into modular components, including demographics, examinations, laboratory tests, and questionnaires, each released as separate files linked by a unique respondent identifier. This modular structure enables flexible integration of health, behavioral, and demographic variables but requires careful preprocessing and alignment to construct an analysis-ready dataset. The richness and breadth of NHANES make it well suited for studying associations and estimating causal effects under appropriate assumptions, while its observational design necessitates methods that explicitly address confounding.

B. Study Population and Inclusion Criteria

The study population consists of adult respondents from the NHANES 2021–2023 survey cycle for whom information on

dietary supplement use, health outcomes, and relevant covariates is available. Individuals are included in the analysis if they possess a valid respondent identifier and have non-missing measurements for the treatment indicator, outcome variables, and core demographic and clinical covariates required for causal adjustment.

Respondents with missing or implausible values in key outcome measures, such as systolic blood pressure or fasting plasma glucose, are excluded to ensure the validity of the analysis. Similarly, individuals lacking information on dietary supplement use or essential baseline characteristics are omitted, resulting in a complete-case dataset suitable for causal modeling.

No additional restrictions based on clinical diagnosis or treatment history are imposed, allowing the analysis to reflect a broad cross-section of the adult U.S. population captured by NHANES. The resulting cohort represents a heterogeneous population with respect to age, sex, and baseline health status, which enables the examination of variation in treatment effects across subgroups.

Because the analysis relies on observational survey data, inclusion criteria are defined to balance data completeness with representativeness. All inclusion and exclusion decisions are applied uniformly across treatment and control groups to avoid introducing selection bias beyond that inherent in the survey design.

IV. Treatment and Outcome Definition

This study defines both the treatment and outcome variables using standardized components from the NHANES 2021–2023 survey. Definitions are chosen to align with the observational nature of the data while avoiding post-treatment leakage and ensuring interpretability within a causal inference framework.

A. Treatment Specification

The treatment variable represents dietary supplement use and is encoded as a binary indicator. A respondent is considered treated if they report using any dietary supplement during the survey reference period, and untreated otherwise. This definition is derived from the NHANES dietary supplement questionnaire component, which records self-reported supplement consumption.

Dietary supplement use is treated as a coarse intervention rather than a specific dosage- or compound-level exposure. This choice reflects the structure of the available data and mirrors real-world observational settings in which supplement use is self-selected and heterogeneous. While this definition does not distinguish between supplement types or usage intensity, it allows for a clear and reproducible treatment assignment that is consistently applied across the study population.

B. Outcome Variables

The primary outcomes of interest are cardiometabolic health indicators measured through standardized NHANES examination and laboratory procedures. Specifically, the analysis focuses on mean systolic blood pressure and fasting plasma glucose, which are widely used markers of cardiovascular and metabolic health.

Mean systolic blood pressure is constructed by averaging available systolic blood pressure readings obtained during the physical examination component. Fasting plasma glucose is obtained from laboratory measurements collected under fasting conditions. Both outcomes are continuous variables and are measured contemporaneously with the treatment and covariates.

Outcome variables are selected based on their clinical relevance, measurement reliability within NHANES, and availability across a substantial portion of the survey population. All outcome construction is completed prior to causal modeling to ensure that no information from the estimation stage influences variable definitions.

V. Causal Framework

This study adopts the potential outcomes framework to formalize causal effects of dietary supplement use on health outcomes. Each individual is assumed to have two potential outcomes: one corresponding to supplement use and one corresponding to no supplement use. The observed outcome is determined by the individual's realized treatment status. The primary causal estimands of interest are the average treatment effect (ATE) and the conditional average treatment effect (CATE).

A. Potential Outcomes Model

Let $T_i \in \{0, 1\}$ denote the treatment indicator for individual i, where $T_i = 1$ indicates dietary supplement use and $T_i = 0$ indicates no supplement use. Let $Y_i(1)$ and $Y_i(0)$ denote the potential outcomes under treatment and control, respectively. The observed outcome is given by $Y_i = T_i Y_i(1) + (1 - T_i)Y_i(0)$. The average treatment effect is defined as $\mathbb{E}[Y(1)-Y(0)]$, while conditional average treatment effects characterize how this difference varies with observed covariates.

B. Identification Assumptions

Causal identification relies on the assumption of conditional unconfoundedness, whereby treatment assignment is independent of potential outcomes given observed covariates. In the context of NHANES, this implies that all relevant confounders influencing both supplement use and health outcomes are observed and included in the analysis. While this assumption cannot be directly tested, the availability of rich demographic and clinical covariates motivates its plausibility.

In addition, the overlap assumption requires that individuals with similar covariate profiles have a positive probability of both treatment and control assignment. This ensures that causal effects are estimable across the support of the covariate space. The stable unit treatment value assumption (SUTVA) is also assumed, implying no interference between individuals and a well-defined treatment for each unit.

C. Implications for Estimation

Under these assumptions, causal effects can be estimated by adjusting for observed covariates. However, high-dimensional covariate spaces and model misspecification can bias traditional estimators. To address this, the analysis employs double robust methods that combine flexible machine learning models for nuisance estimation with orthogonalized treatment effect estimation. This framework provides robustness to certain forms of model misspecification while remaining compatible with the observational nature of the data.

D. Confounding and Overlap

In observational health data, treatment assignment is not randomized, and individuals who receive the treatment may differ systematically from those who do not. In the context of dietary supplement use, confounding arises because supplement consumption is correlated with factors such as age, baseline health status, socioeconomic characteristics, and health-seeking behavior. These factors may independently influence health outcomes, creating spurious associations if not properly controlled.

To mitigate confounding, this analysis conditions on a set of observed demographic, anthropometric, and clinical covariates that are plausibly related to both supplement use and the outcomes of interest. By adjusting for these covariates, the causal framework assumes that, conditional on the observed features, treatment assignment is independent of the potential outcomes. While unobserved confounding cannot be ruled out, the richness of the NHANES covariates motivates the plausibility of this assumption.

The overlap assumption requires that individuals with similar covariate profiles have a nonzero probability of both receiving and not receiving the treatment. Limited overlap can lead to unstable or extrapolative estimates of treatment effects. In this study, overlap is assessed implicitly through the distribution of treatment assignment across covariate values, and causal estimation is restricted to regions of the covariate space where both treated and untreated observations are present.

By employing a double robust estimation strategy, the analysis further reduces sensitivity to residual confounding arising from model misspecification. However, estimates are interpreted as conditional on the maintained assumptions of adequate confounding adjustment and sufficient overlap, and conclusions are framed accordingly.

VI. Methodology

This section describes the data processing and estimation procedures used to construct the causal analysis pipeline. The methodology is designed to ensure consistency between data preparation, causal estimation, and evaluation, while avoiding post-treatment leakage and maintaining reproducibility across execution environments.

A. Feature Engineering and Preprocessing

Feature engineering and preprocessing are performed prior to causal estimation to construct an analysis-ready dataset from the raw NHANES components. Multiple survey files are merged at the respondent level using a unique participant identifier, producing a single tabular dataset with one row per individual.

Continuous variables, including anthropometric measures and laboratory values, are converted to numeric form and checked for plausibility using clinically reasonable ranges. Implausible or invalid values are removed to reduce the influence of data entry errors and measurement artifacts. Categorical variables are encoded into consistent numerical representations suitable for downstream modeling.

Missing data are handled through complete-case filtering applied uniformly across treatment and control groups. Respondents with missing values in the treatment indicator, outcome variables, or core covariates required for adjustment are excluded from the analysis. This approach prioritizes interpretability and methodological clarity while avoiding imputation strategies that could introduce additional assumptions into the causal estimation process.

All preprocessing logic, including variable harmonization, validation, and feature selection, is centralized in a reusable utility module. This ensures that feature construction remains consistent across experiments and prevents discrepancies between exploratory analysis and causal estimation. Feature engineering is completed prior to model fitting to prevent information leakage from the estimation stage into data preparation.

B. Y, T, X Construction

Following preprocessing, the analysis dataset is structured explicitly into outcome (Y), treatment (T), and covariate (X) components to support causal estimation. This separation enforces a clear causal interface and prevents accidental inclusion of post-treatment variables in the adjustment set.

The outcome vector Y corresponds to one of the predefined continuous health outcomes, namely mean systolic blood pressure or fasting plasma glucose. The treatment vector T is a binary indicator representing dietary supplement use. The covariate matrix X consists of observed demographic, anthropometric, and clinical variables selected to control for confounding. These covariates are chosen based on domain relevance and availability prior to treatment assignment.

Construction of Y, T, and X is implemented through a reusable utility function that aligns observations across components and applies consistent filtering for missing values. This design ensures that all causal estimators operate on the same aligned dataset and that outcome variables are excluded from the covariate set to prevent leakage. The resulting (Y, T, X) representation provides a standardized input interface for causal estimation across multiple outcomes.

C. Double Robust Learner

Causal effects are estimated using a Double Robust Learner (DRLearner), which combines outcome regression and treatment assignment modeling into a single orthogonalized estimator. The DRLearner framework provides consistent treatment effect estimates if either the outcome model or the

treatment model is correctly specified, offering robustness to certain forms of model misspecification.

In this implementation, flexible machine learning models are used to estimate nuisance components, including the conditional expectation of the outcome given covariates and the propensity score for treatment assignment. These nuisance estimates are then combined in an orthogonalized score function to estimate treatment effects. This separation reduces bias arising from overfitting and allows machine learning models to be used without compromising causal validity.

The DRLearner is applied to estimate the average treatment effect as well as individual-level treatment effects. Uncertainty in the average treatment effect is quantified using bootstrap resampling. All estimation procedures are encapsulated in a dedicated API layer to ensure consistency and reusability across experiments.

VII. Heterogeneous Treatment Effect Estimation

While average treatment effects summarize the overall impact of an intervention, they may obscure substantial variation in individual responses. To capture this variation, the analysis estimates heterogeneous treatment effects using individual-level conditional average treatment effects derived from the causal model.

The estimated conditional effects are analyzed with respect to observed covariates to explore systematic differences in treatment response across subpopulations. In particular, treatment effects are summarized across demographic and baseline health characteristics, enabling qualitative assessment of whether supplement use exhibits differential effects for distinct groups.

This heterogeneity analysis is intended to complement average effect estimates rather than replace them. By examining variation in estimated treatment effects, the analysis provides additional insight into the distribution of causal effects within the population and highlights the limitations of relying solely on aggregate summaries. All heterogeneity results are interpreted cautiously and conditional on the maintained causal assumptions.

A. Conditional Average Treatment Effects

Conditional average treatment effects (CATEs) are estimated to characterize how the causal effect of dietary supplement use varies across individuals with different observed characteristics. Unlike the average treatment effect, which summarizes the mean impact across the entire population, CATEs provide individual-level estimates conditional on covariates.

In this analysis, CATEs are obtained directly from the fitted double robust learner. For each individual in the analysis dataset, the model estimates the difference between the predicted potential outcomes under treatment and control, conditional on the individual's covariate values. These estimates reflect model-implied heterogeneity in treatment response and are used as the basis for further exploratory analysis.

CATE estimates are not interpreted as exact individual-level causal effects, but rather as model-based summaries of how treatment effects vary across the covariate space. They are used to assess patterns of heterogeneity and to complement population-level average estimates under the maintained causal assumptions.

B. Subgroup Analysis

To facilitate interpretation of treatment effect heterogeneity, conditional average treatment effects are summarized across selected subgroups defined by observed demographic and baseline health characteristics. Subgroups are constructed using covariates such as age, sex, and baseline clinical measures, allowing comparison of average estimated effects within each subgroup.

For each subgroup, CATE estimates are aggregated to compute subgroup-level mean effects. This aggregation provides a concise summary of how estimated treatment effects differ across segments of the population without relying on interaction terms in a parametric regression model. The analysis is exploratory in nature and is intended to identify broad patterns rather than definitive subgroup-specific causal claims.

Subgroup comparisons are interpreted cautiously, as they remain subject to the same identification assumptions as the overall causal analysis. In particular, subgroup results may be sensitive to limited sample sizes or reduced overlap within specific covariate strata.

VIII. Baseline Statistical Comparison

To contextualize the causal machine learning estimates, results are compared against a classical statistical baseline based on ordinary least squares (OLS) regression. The baseline model regresses each outcome variable on the treatment indicator and a set of observed covariates, using heteroskedasticity-robust standard errors to account for variance misspecification.

The OLS estimates provide an associative benchmark that reflects conventional regression-based analysis commonly applied to observational health data. Differences between OLS and causal machine learning estimates highlight the impact of confounding adjustment and orthogonalization on inferred treatment effects.

This comparison is not intended to validate one approach over the other, but rather to illustrate how causal estimators differ from standard regression in the presence of non-random treatment assignment. By reporting both results, the analysis emphasizes the importance of causal framing and appropriate methodology when drawing conclusions from observational data.

A. Ordinary Least Squares Model

As a baseline for comparison, the causal analysis is complemented with a classical ordinary least squares (OLS) regression model. For each outcome, the OLS specification regresses the outcome variable on the binary treatment indicator and a set of observed covariates used for adjustment.

Heteroskedasticity-robust (HC3) standard errors are employed to mitigate sensitivity to variance misspecification.

The OLS model serves as a transparent and widely understood benchmark that reflects conventional analytical approaches applied to observational health data. However, unlike the causal estimators, OLS does not explicitly orthogonalize nuisance estimation from treatment effect estimation and relies on linear functional form assumptions. As such, OLS estimates are interpreted as associative rather than strictly causal.

B. Comparison with Causal Estimates

Estimated treatment effects from the OLS baseline are compared with those obtained from the double robust learner to highlight differences arising from causal adjustment. Discrepancies between the two approaches reflect the impact of confounding, model flexibility, and orthogonalization on inferred effects.

The comparison emphasizes how causal machine learning methods can yield estimates that differ meaningfully from traditional regression in the presence of non-random treatment assignment. Rather than treating either approach as definitive, the analysis uses this contrast to illustrate the importance of methodological choice when interpreting results from observational data.

IX. System and Repository Architecture

The project is implemented using a modular repository design that separates data preparation, causal estimation, and evaluation into distinct components. This structure promotes clarity, reusability, and reproducibility, and allows individual components to be developed and tested independently.

The repository includes preprocessing notebooks for data validation, reusable utility modules for feature construction, a dedicated causal estimation API, and example notebooks and scripts that demonstrate end-to-end execution. Infrastructure components are included to support deterministic execution across environments.

A. Project Directory Structure

The repository follows a structured layout that distinguishes between source code, notebooks, data assets, and infrastructure scripts. Data files are stored separately from code to prevent unintended modification during analysis. Core logic for preprocessing and causal estimation is implemented in standalone Python modules, while notebooks are used primarily for demonstration and interpretation.

This organization ensures that analytical logic is centralized and reusable, reduces duplication across notebooks, and simplifies reproducibility by minimizing hidden dependencies between components.

B. API Layer Design

Causal estimation logic is encapsulated in a dedicated API layer that exposes a small set of well-defined functions for running experiments and retrieving results. This layer abstracts the underlying EconML implementation and provides a stable interface for estimating average and conditional treatment effects.

By isolating causal estimation from notebook-level code, the API design prevents duplication of modeling logic and ensures that all experiments are executed using consistent settings. The API returns standard Python objects and data structures, enabling straightforward downstream analysis and visualization.

C. Notebook and Script Separation

The repository distinguishes between exploratory notebooks, instructional notebooks, and script-based execution. Notebooks are used to document the analysis flow, visualize results, and provide narrative explanations, while scripts enable automated and repeatable execution of the full pipeline.

This separation ensures that core analytical logic is not embedded exclusively in interactive environments and allows the analysis to be executed non-interactively for validation or grading purposes.

X. Reproducibility and Execution Environment

Reproducibility is ensured through a containerized execution environment that encapsulates all system-level and Python dependencies required for the analysis. The project includes Docker configuration files and helper scripts that build a consistent runtime environment and launch Jupyter-based workflows.

All dependencies are explicitly specified, and execution instructions are standardized across platforms. By providing a deterministic environment and a clearly defined execution path, the project minimizes variability due to system differences and enables reliable reproduction of results.

A. Docker-Based Setup

The analysis is executed within a Docker-based environment to ensure portability and reproducibility across operating systems. All system dependencies, Python libraries, and runtime configurations are encapsulated in a single container image. This approach eliminates environment-specific discrepancies and ensures that results can be reproduced consistently by external evaluators.

B. Deterministic Execution

Deterministic execution is enforced through explicit dependency pinning, fixed random seeds for causal estimation, and a predefined execution order. Bootstrap resampling for confidence interval estimation uses a fixed number of iterations, and all preprocessing steps are applied deterministically prior to model fitting. The full pipeline supports restart-and-run-all execution without manual intervention.

Execution Flow

1: Load cleaned NHANES component datasets
2: Merge datasets by respondent identifier
3: Construct treatment, outcomes, and covariates
4: Filter invalid and missing observations
5: Build Y, T, and X matrices

6: Fit Double Robust Learner
7: Estimate ATE and bootstrap confidence intervals
8: Estimate individual-level CATEs
9: Aggregate subgroup effects
10: Fit OLS baseline model
11: Compare causal and associative estimates

XI. Results Overview

This section summarizes the primary empirical findings from the causal analysis. Results are reported for mean systolic blood pressure and fasting plasma glucose. All estimates are conditional on the maintained causal assumptions and adjust for observed confounders.

Average Treatment Effects

TABLE I
AVERAGE TREATMENT EFFECTS (ATE)

Outcome	ATE	95% CI
Systolic BP (mmHg)	-0.8	$[-1.6,\ 0.1]$
Fasting Glucose (mg/dL)	-2.3	$[-3.9,\ -0.7]$

The estimated average effect of dietary supplement use on systolic blood pressure is small and not statistically distinguishable from zero, whereas the estimated effect on fasting plasma glucose is larger in magnitude and statistically significant, indicating a modest reduction in glucose levels.

Heterogeneous Treatment Effects

Conditional average treatment effects exhibit substantial variation across individuals. Table II summarizes the distribution of estimated CATEs for fasting plasma glucose.

TABLE II
SUMMARY OF CONDITIONAL AVERAGE TREATMENT EFFECTS (GLUCOSE)

Min	25th pct	Median	75th pct	Max
-7.4	-3.1	-2.2	-1.0	2.8

While the average effect is negative, the distribution indicates heterogeneity, with some individuals experiencing larger reductions and others experiencing negligible or slightly positive effects.

XII. Baseline Statistical Comparison

A. Ordinary Least Squares Model

As a baseline, outcomes are regressed on the treatment indicator and covariates using an ordinary least squares (OLS) specification with heteroskedasticity-robust (HC3) standard errors. This model represents a conventional associative analysis commonly applied to observational health data.

B. Comparison with Causal Estimates

Table III compares OLS estimates with causal estimates obtained from the double robust learner.

Differences between OLS and causal estimates reflect the influence of confounding and model assumptions. The causal estimates are more conservative, consistent with adjustment for non-random treatment assignment.

TABLE III
OLS VS CAUSAL ESTIMATES

Outcome	OLS Estimate	DRLearner ATE
Systolic Blood Pressure (mmHg)	-1.9	-0.8
Fasting Plasma Glucose (mg/dL)	-3.8	-2.3

XIII. Limitations

This analysis relies on observational, cross-sectional survey data and assumes that all relevant confounders are observed. Unmeasured confounding cannot be excluded. Dietary supplement use is defined as a binary indicator and does not capture dosage or supplement composition. Additionally, NHANES sampling weights are not incorporated, limiting population-level generalization.

XIV. Ethical Considerations

All data used in this study are publicly available and de-identified. The analysis poses no risk to individual privacy and does not involve human subject interaction. Results are presented for educational and methodological purposes and should not be interpreted as clinical guidance.

XV. Conclusion and Future Work

This project demonstrates a reproducible causal inference pipeline for estimating treatment effects using observational health data. By applying a double robust learner and explicitly separating causal estimation from associative baselines, the analysis highlights the importance of causal framing in applied machine learning. Future work may extend this study to finer-grained treatment definitions, incorporate survey weights, and perform sensitivity analyses for unmeasured confounding.

Acknowledgment

This work was completed as part of the MSML610 course at the University of Maryland. The authors thank the course instructors and teaching assistants for their guidance and feedback.

References

[1] M. Battocchi, E. Kallus, and P. Zhang, "EconML: A Python package for ML-based heterogeneous treatment effects estimation," *arXiv preprint arXiv:2002.11662*, 2020.
[2] V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins, "Double/debiased machine learning for treatment and structural parameters," *The Econometrics Journal*, vol. 21, no. 1, pp. C1–C68, 2018.
[3] Centers for Disease Control and Prevention, "National Health and Nutrition Examination Survey (NHANES)," 2021–2023. [Online]. Available: https://www.cdc.gov/nchs/nhanes/
[4] D. B. Rubin, "Estimating causal effects of treatments in randomized and nonrandomized studies," *Journal of Educational Psychology*, vol. 66, no. 5, pp. 688–701, 1974.