Medications and conditions associated with weight loss in patients prescribed semaglutide based on real-world data

William Powell, BS[1], **Xing Song, PhD**[2], **Yahia Mohamed, MBChB, MS**[1], **Dave Walsh, BS**[1], **Elizabeth Parks, PhD**[3], **Tamara M McMahon, MA**[1], **Mirza Khan, MD, MS**[4,5], **Lemuel R Waitman, PhD**[1,2]

[1]Department of Biomedical and Health Informatics, University of Missouri-Kansas City School of Medicine, Kansas City, Missouri, USA

[2]Department of Health Management and Informatics, University of Missouri School of Medicine, Columbia, Missouri, USA

[3]Department of Nutrition and Exercise Physiology, University of Missouri, Columbia, Missouri, USA

[4]Department of Cardiovascular Medicine, Saint Luke's Mid America Heart Institute; Kansas City, Missouri, USA

[5]Section of Cardiology, University of Missouri-Kansas City; Kansas City, Missouri, USA.

Abstract

Objective—Approved by the Food and Drug Administration (FDA) in 2017 for diabetes and 2021 for weight loss, semaglutide has seen widespread use among individuals who aim to lose weight. We sought to evaluate weight loss and the influence of clinical factors on semaglutide patients in real-world clinical practice.

Methods—Using data from 10 Health Systems within the Greater Plains Collaborative (a PCORnet Clinical Research Network), we extracted nearly 4,000 clinical factors encompassing demographic, diagnosis, and prescription information for semaglutide patients. A gradient boosting machine learning classifier was developed for weight loss prediction and identification of the most impactful factors via SHAP (SHapley Additive exPlanations) value extrapolation.

Results—We studied 3,555 eligible patients (539 of whom were observed 52 weeks following exposure) from March 2017 to April 2022. On average, individuals lost 4.44% (Males 3.66%, Females 5.08%) of their initial weight. History of diabetes mellitus diagnosis was associated with less weight loss while prediabetes and linaclotide use were associated with more pronounced weight loss.

Conclusion—Weight loss in patients prescribed semaglutide from real-world evidence was strong but attenuated compared to previous clinical trials. Machine learning analysis of electronic

---

Contact Info: Russ Waitman, PhD, russ.waitman@health.missouri.edu, Address: University of Missouri School of Medicine 1 Hospital Dr., Columbia, MO 65212.

health record data identified factors that warrant further research and consideration when tailoring weight loss therapy.

Keywords

Glucagon-Like Peptide 1 (GLP-1); Weight-Reducing Drugs; Databases

Introduction

Obesity rates in the United States remain high (1) and are associated with numerous health conditions such as diabetes, heart disease, stroke, and higher mortality rates (2). While interventions for individuals with obesity include changes in behavior patterns, such as diet and exercise (3), adoption is challenging and often not associated with clinically meaningful weight loss (4); further motivating the development of novel prescription drugs for weight loss.

Semaglutide, a Food and Drug Administration (FDA)-approved medication intended to treat type 2 diabetes mellitus or reduce the risk of heart disease in such patients (5), has shown promise as a weight loss drug. In a randomized controlled trial, those on semaglutide experienced an average body weight reduction of up to 13% and 10% greater decrease in body weight on average than those on placebo (6). Additionally, after 52 weeks, up to 65% of patients lost at least 10% in body weight (6).

Semaglutide's promise from controlled studies motivated our study to examine real-world weight loss across multiple healthcare systems and explore factors that are associated with improved or reduced effectiveness. We analyzed electronic health record (EHR) data, using the Greater Plains Collaborative (GPC) (7), a PCORnet Clinical Research Network (8). We describe the composition and magnitude of weight loss in patients prescribed semaglutide and then apply machine leaning models to gain insight the association of other prescription medications, diagnoses, and demographic information relative to the impact of semaglutide on weight loss.

Methods

Greater Plains Collaborative EHR Records

Annually, EHR and billing information from thirteen GPC healthcare systems ("sites") are integrated in the Greater Plains Collaborative Reusable Observable Unified Study Environment (GROUSE) (9) which creates interoperable deidentified databases using the PCORnet Common Data Model (CDM) (10) format. These records include patient demographic, prescription, and diagnosis information used in this research. Three healthcare systems did not have adequate information or data conformance for semaglutide patients that met the study criteria and were excluded; as a result 10 health systems with observations from March 2017 to April 2022 were used. All protocols were approved by the University of Missouri Institutional Review Board.

Study Criteria: Semaglutide Exposure and Weight Observation

The present study examined individuals that were prescribed semaglutide to analyze its overall effect on weight loss and determine the impact of relevant clinical factors on this loss. The semaglutide exposure focused on individuals prescribed dosages of 0.25mg to 2mg per week in line with Phase 2 of the clinical trial for semaglutide (6) as diabetes therapy or off label for weight loss. Individuals were included with underlying demographic information available as well as longitudinal data on semaglutide exposure and weight measurement. Using machine learning techniques, we analyzed clinically relevant demographic, diagnosis, and prescription factors and their relationship to weight loss performance of patients tracked 52 weeks after starting semaglutide.

To allow for observation of clinically meaningful effects of treatment, semaglutide patients were only included in the study if they had at least 3 recorded prescriptions and duration of semaglutide exposure of at least 12 weeks (Figure 1). Weight measurements were tracked relative to the window of semaglutide exposure. Initial weight was the measurement most recently occurring in a 30-day window leading up to and on the day of obtaining first semaglutide prescription. For overall observation the final weight was the most recent weight observation that would have occurred any amount of time at least 12 weeks after taking semaglutide. For machine learning analysis focused on clinical factors associated with weight loss, the end weight was measured one year (50–52 weeks) after beginning taking semaglutide. Individuals were required to have at least 26 weeks of semaglutide prescription out of this yearlong observation period. Weight change was classified as the percent change from the initial to the final weight measurements and categorized into a binary outcome with successful weight loss defined as 10% reduction in weight.

Prescription Medication, Diagnosis, and Demographic Information

Other medications for the semaglutide cohort focused on prescriptions written around the time of semaglutide exposure and defined as all prescriptions occurring in the 365 days leading up to the first semaglutide exposure through the date of their final weight. Prescribed medications were grouped using ingredient level RxNorm (11) Concept-unique identifier (CUI) codes obtained via the RxNorm getDrugs Application Programming Interface (API) (12). Medication analysis found 801 unique groups of drug ingredients prescribed to semaglutide patients during the observation period from a year before semaglutide exposure to the end date of a year after semaglutide exposure, shown in the "prescription observation window" shown in Figure 1.

Patient diagnoses recorded as International Classification of Diseases, Ninth Revision (ICD-9) or Tenth Revision (ICD-10) codes (13) in sites' EHR and billing systems were mapped to diagnosis groups using Phenotype Codes (Phecodes) (14,15). Phecodes group ICD codes into clinically meaningful groups. Patients were excluded if they had 1) an active cancer diagnosis in the past year (ICD-9 140–209.99, 230–240, ICD-10 C&D) except for benign neoplasms (ICD-9 210–229.99, ICD-10 D10-D36), 2) a history of bariatric surgery, 3) being underweight, or 4) pregnancy. Diagnoses only included those in the year prior to beginning semaglutide through the window of semaglutide exposure.

Demographic and vital sign information was obtained and included the patient's sex, race, age at start of semaglutide prescriptions, and initial weight at the time of semaglutide initiation. The duration of the semaglutide prescription was recorded in weeks. Sex and race were recorded as binary variables with male and female for sex and white and non-white for race, respectively. Patients with multiple birthdays were excluded.

Semaglutide Weight-Loss in Real-World Settings

We calculated percent weight loss for patients prescribed semaglutide across the total population and stratified by sex. The mean and standard deviation were calculated along with density plots to visualize the distribution of percent weight loss for men and women. In addition, we analyzed the percent of patients that achieved 5% and 10% weight loss, overall and stratified by sex. This process was repeated for weight loss in the maximally observable window, as well as for weight loss at 12 weeks and 52 weeks after first semaglutide exposure.

We also describe a 95% confidence interval for the population mean percent change in weight demonstrating the plausible range of values for the proportion of semaglutide patients that lost 5% or 10% of weight. These metrics allowed for analysis of mean weight loss and likelihood of patients achieving 5% or 10% weight loss on semaglutide overall, by sex, and by duration of exposure. We also reported on the difference in proportions between males and females that lost 5% or 10% in weight while on semaglutide.

Predictive Modeling and Risk Factor Discovery

We adopted a gradient boosting machine (GBM) model, an embedded feature selection technique which performed feature selection while constructing and optimizing a prediction model, on the two subgroups separately (16). GBM is an ensemble learning technique that generates a sequence of decision trees, each of which is designed to further improve prediction accuracy from the previous trees. We developed and validated the model with 5-fold cross-validation (17) and repeated the experiment 10 times to evaluate result stability. Predictions were evaluated based on area under the receiver operator characteristic curve (AUROC), specificity, sensitivity, and precision on testing datasets. To control for overfitting, we carefully tuned the model hyper-parameters (i.e., depth, learning rate, number of iterations, L2 regularization term, random strength, and bagging temperature) within each training session. We compared between two state-of-art GBM implementations and adopted Catboost to generate the final predictive model and risk factor set, as it demonstrated superior performance in AUROC over other implementations (e.g., xgboost (18)) in presence of predominantly categorical features. Catboost (19,20) is a novel gradient boosting toolkit tailored for categorical variables that can be used for classification or regression, here used for classification.

A total of 1,395 predictors were initially included in the Catboost model, including: 588 Phecodes; 801 medication ingredients that were present in the population out of 3,102 possible medication groups; basic demographic information such as age, sex, and race; time (in weeks) on semaglutide; as well as initial weight. We evaluated the marginal effects of

each feature using the Shapley Additive exPlanations (SHAP) value (21), which measured how the predicted odds ratio would change by including a particular factor of certain value for each individual patient (22). The SHAP value extrapolations were fit with the publicly available SHAP API (23), Tree Explainer (24) and Force plots (25). Feature importance was ranked based on average SHAP values. Using the top 20 important features from our SHAP value analysis, we fit a parsimonious logistic regression model and reported adjusted odds ratio (OR). All analyses were conducted in Python 3.7.017 using open source packages (26–32).

Results

Study population

Of the 36,318 individuals across 13 health systems exposed to semaglutide, 11,349 from the 10 health systems included in our analysis were tracked over 12 weeks. After screening for individuals that also had demographic information available, trackable weight records in the required time windows, and did not meet any exclusion criteria of cancer, bariatric surgery, or pregnancy, 3,555 patients remained for study. These individuals were used to assess overall weight change associated with semaglutide exposure. To examine associated 52 week weight change, we also conducted a subgroup analysis of patients exposed to semaglutide for at least 26 weeks with weight measurements 52 weeks from semaglutide initiation (n = 539; 15.2% of total semaglutide exposure population). This subpopulation was also studied using machine learning analysis on other clinically relevant factors for weight change in combination with semaglutide (Figure 2). The study population was primarily white, with 2606 (74.4%) in the overall population and 384 (71.2%) in the population analyzed after 52 weeks. The average age was 55.0 years and weight was 108.7 kilograms in the overall cohort, while the average age was 55.8 years and weight 107.2 kilograms in the cohort observed over 52 weeks. Diabetes mellitus incidence was seen in 3115 (87.6%) of the total semaglutide population and in 450 (83.4%) of the cohort tracked over 52 weeks (Table 1). Further information on the most common Phecode group coverage in the cohort observed over 52 weeks is shown in Table S1.

Weight Loss Experienced by Individuals on Semaglutide

Individuals on semaglutide experienced weight loss success, and the weight loss at different exposure durations are shown in Table 2. In the overall semaglutide exposure population, patients on average lost 4.44% (95% CI: 4.13%, 4.75%) of their initial body weight, with males losing less at 3.66% (95% CI: 3.25%, 4.07%), and females losing more at 5.08% (95% CI: 4.58%, 5.57%) as shown in Figure 2. In addition, the proportion of patients losing 5% weight was 41.9% (95% CI: 39.8%, 44.0%), and 10% reduction weight was 18.1% (95% CI: 16.4%, 19.8%). At 52 weeks from the start of semaglutide treatment, patients on average lost 4.43% (95% CI: 3.87%, 4.98%) of their initial body weight, with males losing 3.83% (95% CI: 3.11%, 4.55%) and females losing 4.86% (95% CI: 4.06%, 5.66%). Individuals with a history of Diabetes mellitus lost 7.44% (95% CI: 5.73%, 9.15%) of initial weight after 52 weeks while those without diabetes lost only 3.86% (95% CI: 3.26%, 4.47%). On average, we found greater weight loss was associated with a longer duration of semaglutide exposure, though in diverse real world populations the average loss appeared

to be around 5%. Men and women experienced approximately 3–5% mean weight loss and women experienced greater magnitudes of weight loss compared to men (Figure 3).

Catboost Model Fitting and SHAP Interpretation

The Catboost Classifier model was trained on the data with 10% weight loss as the primary outcome, and obtained an AUROC of 0.808 (95% CI: 0.694, 0.901). Model sensitivity and specificity was 0.217. and 0.982, respectively. Overall accuracy was 0.852 on all data. Given that the model was utilized for feature analysis, its low sensitivity was deemed acceptable and with high AUROC, specificity, and accuracy the analysis proceeded. The best Catboost model achieving the highest AUROC score had hyperparameters with 50 iterations, a learning rate of 0.3, a max depth of 3, and a L2 regularization term of 0.5. The ROC curves for the Catboost classifier and other gradient boosting machine models are shown in Figures S1–S3.

The Average SHAP value plots after 10-fold cross-validations are presented in Figure 4. From this analysis, the factors found most highly associated with weight loss included disorders of the adrenal glands, linaclotide use, elevated blood glucose level, and codeine use. Elevated blood glucose level diagnoses related to prediabetes rather than diabetes mellitus. Diagnosis of diabetes mellitus and history of dulaglutide and metformin use were most associated with limited weight loss success. SHAP feature analysis revealed a combination of demographic, diagnosis, and prescription factors as important features in predicting   10% weight loss (Table S2).

Adjusted odds ratios of achieving 10% weight loss 52 weeks after semaglutide exposure (Table 3) were calculated using coefficients from a logistic regression model, where the logistic regression model was fit using the top 20 features from the SHAP plots. SHAP features found with an adjusted association with weight loss had odds ratios and 95% confidence intervals greater than 1. These features included disorders of the adrenal glands, linaclotide use, prediabetes, and codeine use. A history of diabetes mellitus and dulaglutide prescription were most strongly associated with less weight loss.

Discussion

Observed Weight Loss

While individuals taking semaglutide experienced successful weight loss, achieving clinically meaningful weight loss in a real-world setting remains a challenge. With short exposure times, the proportion of patients achieving clinically significant weight loss was low. At 12 weeks, only 28.95% achieved a 5% weight reduction and 4.91% achieved a 10% weight reduction. At 52 weeks, these proportions increased to 40.07% and 17.25%, respectively. In addition, observed mean weight change was found to be 3.06% at 12 weeks, much different than previously reported, but consistent, if slightly attenuated, with other findings after 3 months of exposure (33), albeit at a greater dosage level. We found that women experienced greater weight loss than men with semaglutide exposure, with the majority of both men and women being exposed to the 0.25mg to 2mg weekly dose of semaglutide. While it is possible that this is due to women weighing less to begin with,

and therefore potentially receiving proportionally greater dose exposure, there may be other factors at play. For one, differences in body composition including muscle mass vs fat levels in males and females could play a role in weight loss performance. Our study population had a higher proportion of non-White patients among females than males, and fewer women treated with semaglutide had underlying diabetes mellitus. Interestingly, patients without diabetes experienced greater weight reduction compared to those with diabetes, which suggests that a higher proportion of women may have been taking semaglutide primarily for weight loss. Furthermore, women included in this study had fewer comorbidities such as diabetes, hypertension, and hyperlipidemia. These findings suggest that further research is needed to fully understand the differences in weight loss response between men and women with semaglutide exposure.

Our findings of less pronounced weight reduction compared to previous findings may be due to several factors. Individuals in randomized controlled studies received counseling and attempted lifestyle changes in addition to semaglutide treatment, while in our study it was possible that individuals did not significantly change their lifestyle while taking semaglutide. The attenuated weight loss performance of semaglutide alone may speak to the importance of implementing these lifestyle interventions concurrently with semaglutide prescription. In addition, adherence was much more strongly confirmed in previous studies (6, 33–35), while our study relied only on EHR records of prescription as measure of semaglutide use. Additionally, some previous studies had majority female participants (33–35). The near equivalent number of male and female patients in our study may have contributed to lower overall weight reduction. This study also focused on individuals with a 0.25mg to 2mg weekly dose of semaglutide. The dosage used in the clinical trials to treat obesity are higher 1.7mg or 2.4mg doses (33). We did not include records of the 2.4mg dosage in the present study due to inadequate accumulation of records from FDA approval in 2021 to the end of our study window in March 2022, which could have contributed to more modest weight loss performance of individuals taking semaglutide. Earlier trials also excluded individuals with diabetes, and individuals taking similar dosages of semaglutide as in our study saw weight loss after 52 weeks between 6% and 13% (6), which is very comparable to the 7.4% of loss observed in individuals without diabetes this study. However, 83.4% of our population consisted of individuals with diabetes mellitus, who saw more modest weight loss of 3.9% of their initial body weight.

Clinically Relevant Factors

Our study highlights the ability of Catboost models to predict weight loss at the 10% threshold, while the SHAP analysis found trends in the feature importance not previously reported. We found that factors significantly associated with more weight loss were disorders of the adrenal glands, prescription of linaclotide, prediabetes, other disorders of the eye, malnutrition and underweight diagnosis, lesions of the mouth, and codeine use. Diabetes mellitus, dulaglutide use, metformin use, higher age, and disorders of the retina were associated with less weight loss.

Prior work suggests that individuals with diabetes have more difficulty losing weight (36), which may explain why diabetes mellitus, metformin, or dulaglutide prescription are

strongly associated with less weight loss on semaglutide. Metformin is a gold standard first line care medication for individuals with diabetes (37), and dulaglutide is another common diabetes medication (38). Individuals taking metformin in conjunction with semaglutide found limited success in weight reduction, which may because they were struggling with diabetes care on metformin alone and were prescribed a second medication in semaglutide. Dulaglutide is also a GLP-1 Receptor Agonist. It may be that individuals switched between or took both drugs in the hopes of seeing additive effects, and failed to see them in weight loss. Further research is warranted to better understand the effect of metformin and dulaglutide use on weight loss in individuals taking semaglutide.

The phenotype group for elevated blood glucose levels encompassed those with prediabetes, but not Type 1 or Type 2 diabetes mellitus. We found that weight loss with semaglutide was greater in individuals with prediabetes compared to those with diabetes mellitus. These findings suggest that the weight loss benefits of semaglutide may be greatest in those with early evidence of diabetes or prediabetes. This may be due, at least in part, to the fact that individuals with prediabetes still have more preserved pancreatic beta cell function and insulin reserve, which may allow for greater weight loss response to semaglutide. These findings have important implications for the treatment of individuals with diabetes and prediabetes, and suggest that further research is needed to fully understand the mechanisms underlying the observed differences in weight loss response. Overall, our study provides important insights into the potential use of semaglutide as a weight loss medication for individuals with diabetes and prediabetes.

Additionally, patients with a Phecode for 'disorders of the retina,' which includes diabetic retinopathy, had less weight loss with semaglutide treatment. Interestingly, those with the Phecode for 'other disorders of the eye' tended to have a more robust weight loss response. Based on our analysis, we suspect that non-specific eye condition diagnostic codes were ordered to associate with annual retinopathy screening among patients with diabetes, particularly for those without a history of diabetic retinopathy. This is in line with our earlier findings, where those with prediabetes or new-onset and well-controlled diabetes had a better weight loss response to semaglutide. The presence of diabetic retinopathy is suggestive of long-standing or poorly controlled diabetes, which may have a less robust weight loss response to semaglutide. Similarly, lesions of the mouth could be present as a significant feature in conjunction with the impact of diabetes mellitus on oral health (39). These findings highlight the importance of further research and consideration of patient characteristics and comorbidities when prescribing semaglutide for weight loss.

Adrenal insufficiency is previously known to be associated with weight loss (40), making its appearance as an influencing factor unsurprising. Linaclotide has been used off-label for weight loss in individuals with obesity or eating disorders (41), showing its precedence as a weight loss mechanism and the potential added effects towards weight reduction when taken with semaglutide. Malnutrition and underweight BMI classification was also found to be associated with more weight loss, and was likely an association with the weight loss itself. Weight loss has been shown as a side effect of codeine exposure and withdrawal in mice (42), thus the effect of codeine on weight loss is potentially additive in combination with semaglutide. Older age has been shown to increase the difficulty of losing weight (43),

and our semaglutide population found the same association with older individuals. Previous trials had populations of an average age of 47 years (6) compared to our population's average age of 55.8 years.

Limitations and Agenda for Future Research

There are a number of limitations in the study. Firstly, we obtain information on semaglutide exposures based on their prescriptions, and do not confirm if actual doses are being administered, leading perhaps to potential lack of adherence issues. Also, we do not have information about the diet and exercise patterns of the individuals under examination, which are often first line recommendations for weight loss care. Our findings of lower weight reduction than reported in clinical trials may be attributed, at least in part, to the fact that counseling on lifestyle modifications, which is routine in clinical trials, may be less common in real-world clinical practice. Of note as well, the model achieved very high specificity (0.982) but poor sensitivity (0.217). This suggests our Catboost model is finding factors that are associated with attenuation more completely than promotion of weight loss. Given that the model lacks strong sensitivity performance, it is possible that some findings were false negative. The model could be finding features that are not as strongly negatively associated with weight loss as indicated. As noted above, our study did not have access to information on individuals that took novel higher dosage levels of semaglutide. The highest level of dosage that is included in our study is 2mg per week, while the semaglutide can be prescribed as high as 2.4mg per week. Future research could implement observational analysis in the real world setting to look at the effect of dosage levels on weight loss in individuals with or without diabetes.

Conclusion

Evidence from multi-site EHR data suggests that patients prescribed semaglutide experience weight loss. Yet, these real-world findings suggest smaller reductions in weight of around 5% compared to early clinical trials with around 10% reductions over the same dosage and exposure time. This highlights the challenges of realistically achieving significant weight loss in the real world compared to the clinical trial setting. Machine learning can serve as a valuable tool for analyzing the complex prescription and biological factors that impact weight loss in patients prescribed semaglutide, including the identification of several factors found associated with weight loss performance. These factors include that use of dulaglutide and metformin, and diagnosis of diabetes mellitus, especially severe forms such as those complicated by diabetic retinopathy, were associated with poor weight loss performance. By contrast, use of codeine and linaclotide and diagnosis of prediabetes were associated with greater weight reduction. In addition, the association of female sex on stronger weight loss performance on semaglutide is significant. Individuals with diabetes are more likely to see moderate but not strong weight loss performance at this dosage level of semaglutide, while those without diabetes are likely to see stronger results. These findings suggest patient-level factors worth considering that may influence weight loss performance with semaglutide and directions for further research on how to personalize this method of weight loss therapy.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Collaborators

Great Plains Collaborative

Sravani Chandaka[1], Kelechi (KayCee) Anuforo[1], Lav Patel[1], Daryl Budine[1], Nathan Hensel[1], Siddharth Satyakam[1], Sharla Smith[1], Dennis Ridenour[1], Cheryl Jernigan[1], Carol Early[1], Kyle Stephens[1], Kathy Jurius[1], Kyle Stephens[1], Kathy Jurius[1], Abbey Sidebottom[2], Cassandra Rodgers[2], Hong Zhong[2], Angie Hare[2], Roman Melamed[2], Curtis Anderson[2], Thomas Schouweile[2], Christine Roering[2], Philip Payne[3], Snehil Gupta[3], John Newland[3], Joyce Balls-Berry[3], Janine Parham[3], Evin Fritschle[3], Shanelle Cripps[4], Kirk Knowlton[4], Channing Hansen[4], Erna Serezlic[4], Benjamin Horne[4], Jeff VanWormer[5], Judith Hase[5], Janet Southworth[5], Eric Larose[5], Mary Davis[5], Laurel Hoeth[5], Sandy Strey[5], Brad Taylor[6], Kris Osinski[6], April Haverty[6], Alex Stoddard[6], Sarah Cornell[6], Phoenix Do[6], Lucy Bailey[6], Beth McDonough[6], Betsy Chrischilles[7], Ryan Carnahan[7], Brian Gryzlak[7], Gi-Yung Ryu[7], Katrina Oaklander[7], Pastor Bruce Hanson[7], Brad McDowell[7], Jarrod Field[7], Abu Mosa[8], Sasha Lawson[8], Jim McClay[8], Soliman Islam[8], Vasanthi Mandhadi[8], Kim Kimminau[8], Dennis Ridenour[8], Jeff Ordway[8], Bill Stephens[8], Russ Waitman[8], Deandra Cassone[8], Xiaofan Niu[8], Lori Wilcox[8], Janelle Greening[8], Carol Geary[9], Goutham Viswanathan[9], Jim Svoboda[9], Jim Campbel[9], Frances (Annette) Wolfe[9], Haddy Bah[10], Todd Bjorklund[10], Jackson Barlocker[10], Josh Spuh[10], Louisa Stark[10], Mike Strong[10], Otolose Fahina Tavake-Pas[10], Rachel Hess[10], Jacob Kean[10], Annie Risenmay[10], Olivia Ellsmore[11], Lissa Persson[11], Kayla Torres Morales[11], Sandi Stanford[11], Mahanaz Syed[11], Rae Schofield[11], Meredith Zozus,[11]Brian Shukwit[12], Matthew Decaro[12], Natalia Heredia[12], Charles Miller[12], Alice Robinson[12], Elmer Bernstam[12], Fatima Ashraf[12], Shiby Antony[13], Juliet Fong Zechner[13], Philip Reeder[13], Cindy Kao[13], Kate Wilkinson[13], Tracy Greer[13], Alice Robinson[13], Lindsay Cowel[13]

Affiliation

[1]University of Kansas Medical Center

[2]Allina Health System

[3]Washington University

[4]Intermountain Healthcare

[5]Marshfield Clinic

[6]Medical College of Wisconsin

[7]University of Iowa Healthcare

[8]University of Missouri

[9]University of Nebraska Medical Center

References

1. Imes CC, Burke LE. The Obesity Epidemic: The United States as a Cautionary Tale for the Rest of the World. Curr Epidemiol Rep. 2014 Jun 1;1(2):82–8. [PubMed: 24977112]

2. Flegal KM, Kit BK, Orpana H, Graubard BI. Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis. JAMA. 2013 Jan 2;309(1):71–82. [PubMed: 23280227]

3. Fock KM, Khoo J. Diet and exercise in management of obesity and overweight. J Gastroenterol Hepatol. 2013 Dec;28 Suppl 4:59–63. [PubMed: 24251706]

4. Swift DL, McGee JE, Earnest CP, Carlisle E, Nygard M, Johannsen NM. The Effects of Exercise and Physical Activity on Weight Loss and Maintenance. Prog Cardiovasc Dis. 2018;61(2):206–13. [PubMed: 30003901]

5. Hughes S, Neumiller JJ. Oral Semaglutide. Clin Diabetes. 2020 Jan;38(1):109–11. [PubMed: 31975761]

6. O'Neil PM, Birkenfeld AL, McGowan B, et al. Efficacy and safety of semaglutide compared with liraglutide and placebo for weight loss in patients with obesity: a randomised, double-blind, placebo and active controlled, dose-ranging, phase 2 trial. Lancet. 2018 Aug 25;392(10148):637–49. [PubMed: 30122305]

7. Waitman LR, Aaronson LS, Nadkarni PM, Connolly DW, Campbell JR. The Greater Plains Collaborative: a PCORnet Clinical Research Data Network. J Am Med Inform Assoc. 2014;21(4):637–41. [PubMed: 24778202]

8. Forrest CB, McTigue KM, Hernandez AF, et al. PCORnet® 2020: current state, accomplishments, and future directions. J Clin Epidemiol. 2021 Jan;129:60–7. [PubMed: 33002635]

9. Waitman LR, Song X, Walpitage DL, et al. Enhancing PCORnet Clinical Research Network data completeness by integrating multistate insurance claims with electronic health records in a cloud environment aligned with CMS security and privacy requirements. J Am Med Inform Assoc. 2022 Mar 15;29(4):660–70. [PubMed: 34897506]

10. Common Data Model (CDM) Specification, Version 6.0.

11. Nelson SJ, Zeng K, Kilbourne J, Powell T, Moore R. Normalized names for clinical drugs: RxNorm at 6 years. J Am Med Inform Assoc. 2011;18(4):441–8. [PubMed: 21515544]

12. getDrugs - RxNorm API [Internet]. [cited 2023 Mar 15]. Available from: https://lhncbc.nlm.nih.gov/RxNav/APIs/api-RxNorm.getDrugs.html

13. Geraci JM, Ashton CM, Kuykendall DH, Johnson ML, Wu L. International Classification of Diseases, 9th Revision, Clinical Modification codes in discharge abstracts are poor measures of complication occurrence in medical inpatients. Med Care. 1997 Jun;35(6):589–602. [PubMed: 9191704]

14. Bastarache L Using Phecodes for Research with the Electronic Health Record: From PheWAS to PheRS. Annu Rev Biomed Data Sci. 2021 Jul 20;4:1–19. [PubMed: 34465180]

15. Wu P, Gifford A, Meng X, et al. Mapping ICD-10 and ICD-10-CM Codes to Phecodes: Workflow Development and Initial Evaluation. JMIR Med Inform. 2019 Nov 29;7(4):e14325. [PubMed: 31553307]

16. Friedman JH. Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics. 2001;29(5):1189–232.

17. Scheda R, Diciotti S. Explanations of Machine Learning Models in Repeated Nested Cross-Validation: An Application in Age Prediction Using Brain Complexity Features. Applied Sciences. 2022 Jan;12(13):6681.

18. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]. San Francisco California USA: ACM; 2016 [cited 2023 Mar 15]. p. 785–94. Available from: 10.1145/2939672.2939785

19. Hancock JT, Khoshgoftaar TM. CatBoost for big data: an interdisciplinary review. Journal of Big Data. 2020 Nov 4;7(1):94. [PubMed: 33169094]

20. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. In: Advances in Neural Information Processing Systems [Internet]. Curran Associates, Inc.; 2018 [cited 2023 Mar 15]. Available from: https://proceedings.neurips.cc/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html

21. Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. In: Advances in Neural Information Processing Systems [Internet]. Curran Associates, Inc; 2017 [cited 2023 Mar 15]. Available from: https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

22. Mitchell R, Frank E, Holmes G. GPUTreeShap: massively parallel exact calculation of SHAP scores for tree ensembles. PeerJ Comput Sci. 2022 Apr 5;8:e880.

23. Welcome to the SHAP documentation — SHAP latest documentation [Internet]. [cited 2023 Mar 15]. Available from: https://shap.readthedocs.io/en/latest/

24. Lundberg SM, Erion G, Chen H, et al. From Local Explanations to Global Understanding with Explainable AI for Trees. Nat Mach Intell. 2020 Jan;2(1):56–67. [PubMed: 32607472]

25. Lundberg SM, Nair B, Vavilala MS, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng. 2018 Oct;2(10):749–60. [PubMed: 31001455]

26. Python Release Python 3.7.0 [Internet]. Python.org. [cited 2023 Mar 15]. Available from: https://www.python.org/downloads/release/python-370/

27. Hunter JD. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering. 2007 May 1;9(03):90–5.

28. McKinney W Data Structures for Statistical Computing in Python. In Austin, Texas; 2010 [cited 2023 Mar 15]. p. 56–61. Available from: https://conference.scipy.org/proceedings/scipy2010/mckinney.html

29. CatBoost [Internet]. [cited 2023 Mar 15]. Available from: https://catboost.ai/en/docs/

30. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. MACHINE LEARNING IN PYTHON.

31. Harris CR, Millman KJ, van der Walt SJ, et al. Array programming with NumPy. Nature. 2020 Sep;585(7825):357–62. [PubMed: 32939066]

32. eXtreme Gradient Boosting [Internet]. Distributed (Deep) Machine Learning Community; 2023 [cited 2023 Mar 15]. Available from: https://github.com/dmlc/xgboost

33. Ghusn W, De la Rosa A, Sacoto D, et al. Weight Loss Outcomes Associated With Semaglutide Treatment for Patients With Overweight or Obesity. JAMA Netw Open. 2022 Sep 1;5(9):e2231982. [PubMed: 36121652]

34. Wilding JPH, Batterham RL, Calanna S, et al. Once-Weekly Semaglutide in Adults with Overweight or Obesity. N Engl J Med. 2021 Mar 18;384(11):989–1002. [PubMed: 33567185]

35. Garvey WT, Batterham RL, Bhatta M, et al. Two-year effects of semaglutide in adults with overweight or obesity: the STEP 5 trial. Nat Med. 2022 Oct;28(10):2083–91. [PubMed: 36216945]

36. Franz MJ. Weight Management: Obesity to Diabetes. Diabetes Spectr. 2017 Aug;30(3):149–53. [PubMed: 28848305]

37. Scarpello JHB, Howlett HCS. Metformin therapy and clinical uses. Diab Vasc Dis Res. 2008 Sep;5(3):157–67. [PubMed: 18777488]

38. Scott LJ. Dulaglutide: A Review in Type 2 Diabetes. Drugs. 2020 Feb;80(2):197–208. [PubMed: 32002850]

39. Rohani B Oral manifestations in patients with diabetes mellitus. World J Diabetes. 2019 Sep 15;10(9):485–9. [PubMed: 31558983]

40. Husebye ES, Pearce SH, Krone NP, Kämpe O. Adrenal insufficiency. Lancet. 2021 Feb 13;397(10274):613–29. [PubMed: 33484633]

41. Cid-Ruzafa J, Lacy BE, Schultze A, et al. Linaclotide utilization and potential for off-label use and misuse in three European countries. Therap Adv Gastroenterol. 2022;15:17562848221100946.

42. Suzuki T, Shimada M, Yoshii T, Yanaura S. Induction of physical dependence on codeine in the rat by drug-admixed food ingestion. Jpn J Pharmacol. 1984 Apr;34(4):441–6. [PubMed: 6539401]

43. Jura M, Kozak LP. Obesity and related consequences to ageing. Age (Dordr). 2016 Feb;38(1):23. [PubMed: 26846415]

Study Importance Questions

- What is already known about this subject?

  – Semaglutide, an anti-diabetic medication approved for long-term weight management, has shown effectiveness in clinical trials, including up to 60% of patients losing at least 10% in weight 52 weeks after prescription.

  – Interacting factors, such as comorbidities and other medications can impact weight loss, and their role on semaglutide's real-world performance for weight loss has not been widely explored.

- What are the new findings in your manuscript?

  – Real-world data from patients prescribed semaglutide suggests strong but attenuated weight loss after 52 weeks compared to clinical trial findings.

  – Interactions with important medications and biological factors can both be associated with deterring and promoting weight loss in patients prescribed semaglutide.

- How might your results change the direction of research or the focus of clinical practice?

  – Identified clinical factors, including diabetes and prescribed medications, may warrant future research regarding mechanisms of action in populations where polypharmacy and multiple comorbidities are common.

  – As real-world data analyses are replicated, stable models incorporating concomitant medications and comorbidities may provide improved, personalized clinical decision support for weight management.
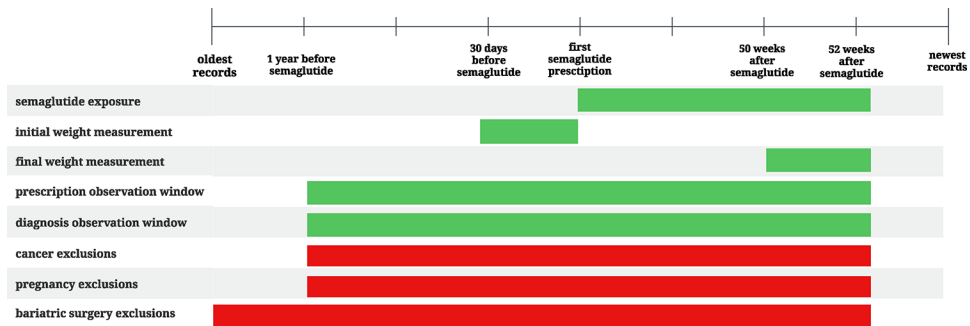
Figure 1.

Data Collection Timeline

The observation period for inclusion and exclusion criteria of the study are shown. Semaglutide prescriptions must have lasted for at least 26 weeks of the year-long observation period, and individuals were required to have a recorded weight in the 30 days leading up to beginning semaglutide and in the period 50–52 weeks after beginning semaglutide. Prescription records and diagnoses were included and individuals with cancer diagnoses were excluded if they occurred in the year prior to the beginning of through the end of semaglutide exposure. Individuals were excluded if they had a record of a pregnancy or bariatric surgery at any point.
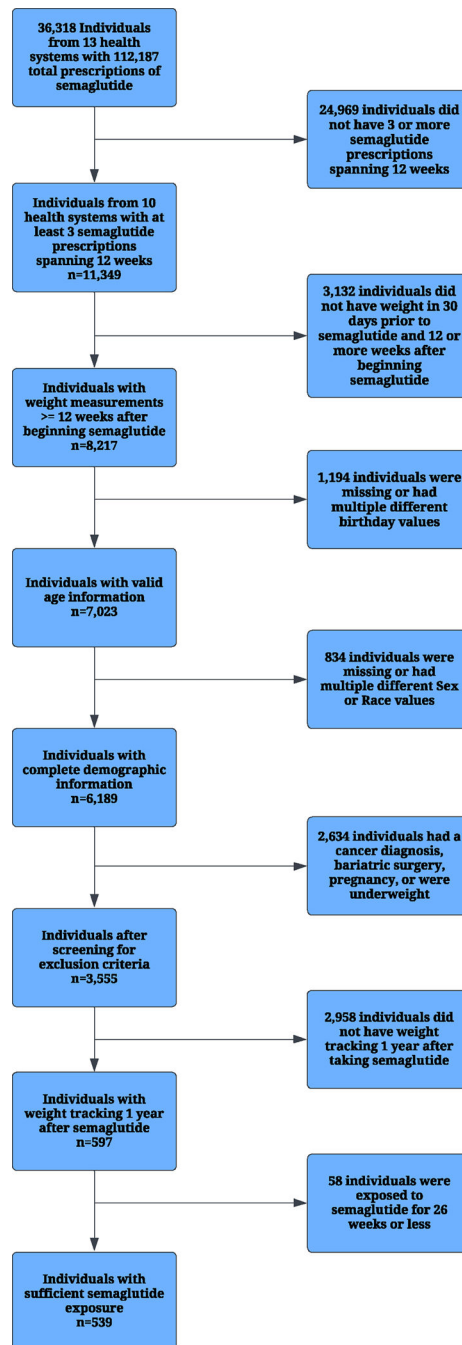
Figure 2.

Patient Cohort Flowchart

Over 36,000 individuals from 13 health systems had an exposure to semaglutide. Following exclusion and complete records criteria the cohort contained 3,555 individuals observable for weight loss. 539 individuals tracked 52 weeks after exposure to semaglutide were used for machine learning analysis.
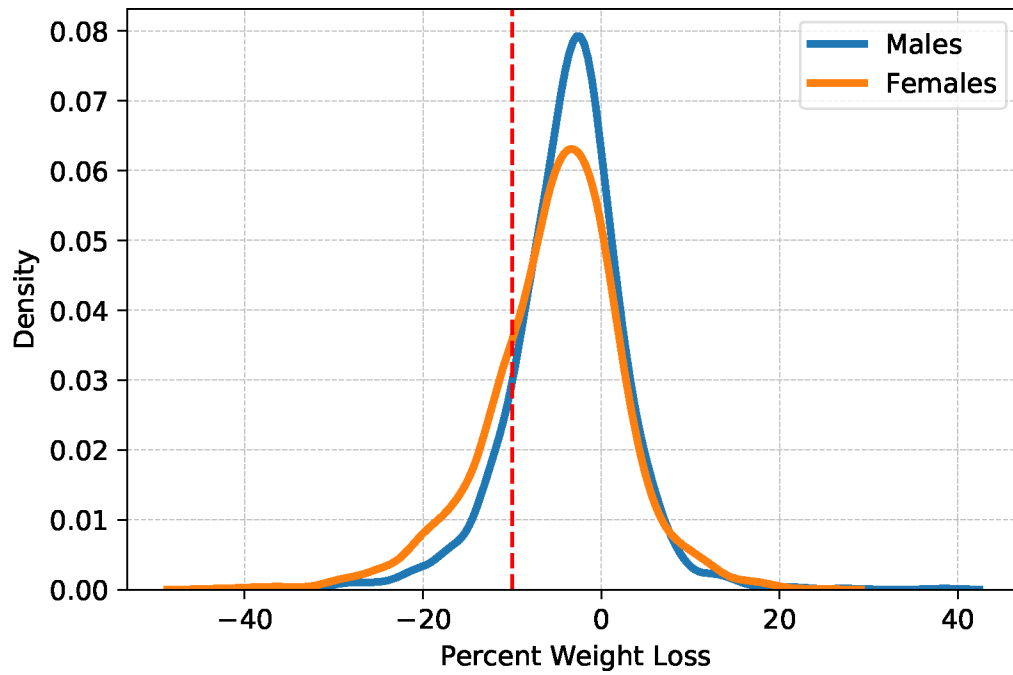
Figure 3.

Density Plot of Weight Loss

Density plots for percent weight loss of semaglutide patients are shown for the 3,555 total trackable individuals on semaglutide, stratified by male and female patients. The threshold of 10% weight loss is shown by the vertical red line. Females experienced a greater proportion of the population losing higher percent of weight than males.
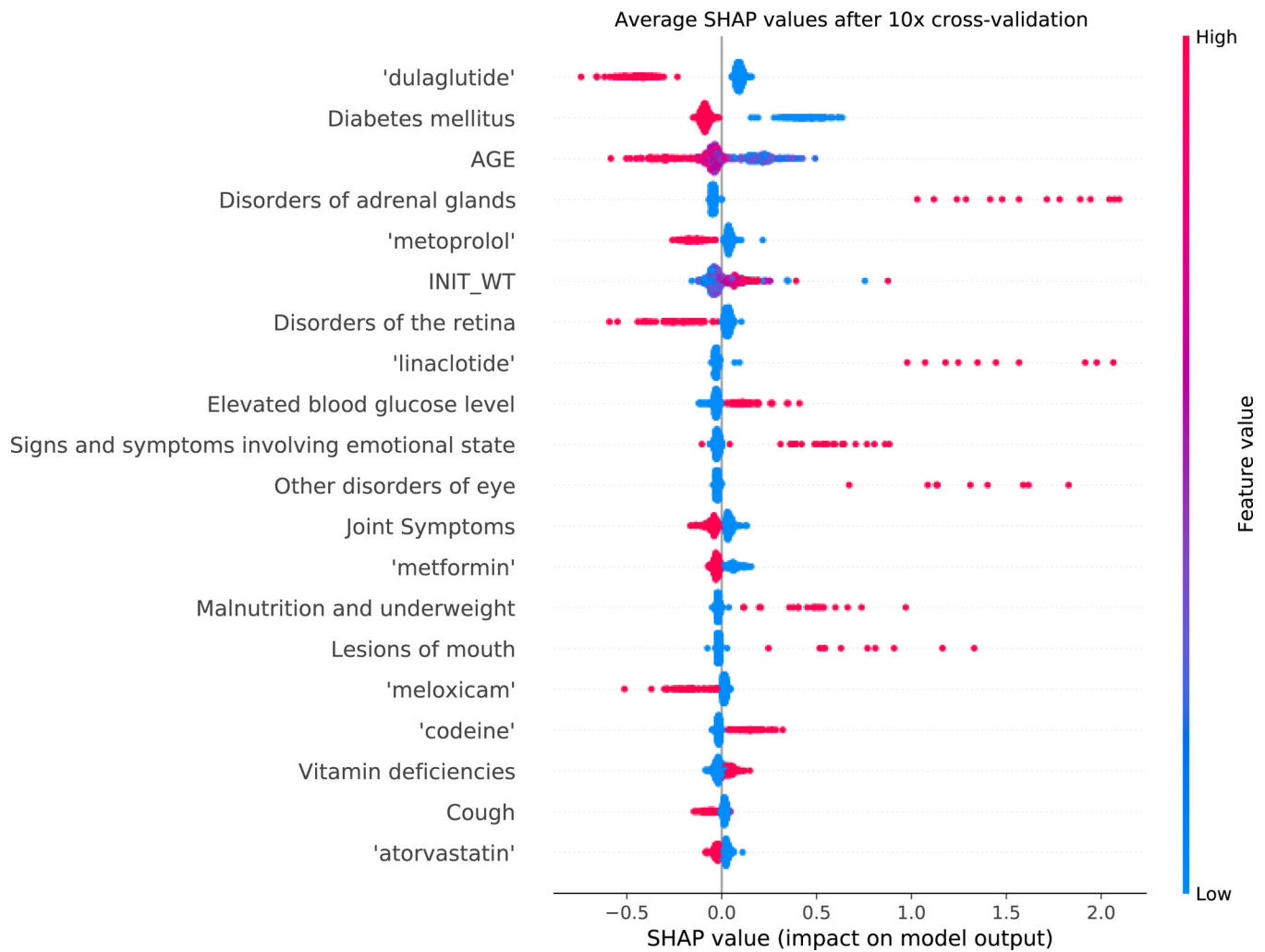
Figure 4.

Repeated 10x Cross-Validation Average SHAP Summary Plot

SHAP value plots for features (27) are presented and ranked vertically by importance. Each individual in the semaglutide study is represented by a dot on each feature row. The dot's position on the horizontal x-axis line represents its SHAP value for that individual, with the dots piling up to show density. Color is used to display the original value of a feature, for example for prescription groups a red dot is used to indicate an individual that has taken that medication (29). Values of the left of zero on the x-axis represent values for that feature influencing the model towards predicting 0, or having not achieved 10% weight loss, while values to the right represent values for the feature influencing the model towards predicting 1, or predicting having successfully achieved 10% weight loss. For example, red values for linaclotide appearing to the right of zero suggest that a patient taking linaclotide is influencing the model to predict that patient taking linaclotide will more likely achieve 10% weight loss on semaglutide. Males are indicated by red coloring and appear to the left, suggesting their relationship is with attenuated weight loss. The global importance of a feature is represented by which features are higher vertically on the plot and are calculated by the mean absolute SHAP value for that feature over all the given samples (29). In the

figure, sex, length of semaglutide exposure, among other features are the most important SHAP features. Complete interpretations of all SHAP features are shown in Table S2. SHAP plots for the XGBoost and Gradient Boost models are presented in Figure S4–S5.

Table 1.

Semaglutide population characteristics

Individual Characteristics	Overall, N = 3,555	Females, N = 1,955	Males, N = 1600	Overall at 52 Weeks, N = 539	Females at 52 Weeks, N = 314	Males at 52 Weeks, N = 225
Race						
White, N (%)	2606 (74.4%)	1353 (69.2%)	1253 (78.3%)	384 (71.2%)	202 (64.3%)	182 (80.9%)
American Indian or Alaska Native, N (%)	25 (0.7%)	17 (0.9%)	8 (0.5%)	6 (1.1%)	6 (1.9%)	0 (0.0%)
Asian, N (%)	91 (2.6%)	51 (2.6%)	40 (2.5%)	18 (3.3%)	11 (3.5%)	7 (3.1%)
Black or African American, N (%)	519 (14.6%)	360 (18.4%)	159 (9.9%)	80 (14.8%)	62 (19.7%)	18 (8.0%)
Native Hawaiian or Other Pacific Islander, N (%)	10 (0.3%)	4 (0.2%)	6 (0.4%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Multiple Race	6 (0.2%)	3 (0.2%)	3 (0.2%)	2 (0.4%)	1 (0.3%)	1 (0.4%)
Refuse to Answer	31 (0.9%)	12 (0.6%)	19 (1.2%)	2 (0.4%)	1 (0.3%)	1 (0.4%)
No Information	12 (0.3%)	9 (0.5%)	3 (0.2%)	2 (0.4%)	2 (0.6%%)	0 (0.0%)
Unknown	151 (4.2%)	90 (5.0%)	61 (3.8%)	21 (3.9%)	12 (3.8%)	9 (4.0%)
Other	122 (3.4%)	74 (3.8%)	48 (3.0%)	31 (5.8%)	21 (6.7%)	10 (4.4%)
Age at start of semaglutide, Mean (SD)	55.0 (12.6)	54.2 (13.1)	56.0 (11.9)	55.8 (12.1)	54.8 (12.6)	57.3 (11.3)
Initial Weight (kg), Mean (SD)	108.7 (26.2)	102.1 (24.1)	116.7 (26.4)	107.2 (25.6)	101.6 (24.0)	114.9 (25.8)
BMI at start of semaglutide, Mean (SD)	37.3 (8.1)	38.0 (8.4)	36.5 (7.6)	37.3 (8.0)	38.0 (8.4)	36.3 (7.2)
Diabetes Mellitus, N (%)	3115 (87.6%)	1616 (82.7%)	1499 (93.7%)	450 (83.4%)	239 (76.1%)	211 (93.8%)
Hyperlipidemia, N (%)	3072 (86.4%)	1624 (83.1%)	1448 (90.5%)	446 (82.7%)	224 (77.7%)	202 (89.8%)
Long Term (Current) Drug Therapy, N (%)	3048 (85.7%)	1675 (85.7%)	1373 (85.8%)	434 (80.5%)	247 (78.7%)	187 (83.1%)
Hypertension, N (%)	2843 (80.0%)	1481 (75.8%)	1362 (85.1%)	415 (77.0%)	225 (71.7%)	190 (84.4%)
metformin, N (%)	2398 (67.4%)	1213 (62.1%)	1185 (74.1%)	403 (67.5%)	200 (63.7%)	167 (74.2%)
atorvastatin, N (%)	1626 (45.7%)	797 (40.8%)	829 (51.8%)	271 (45.4%)	131 (41.7%)	114 (50.7%)
acetaminophen, N (%)	1066 (30.0%)	614 (31.4%)	452 (28.3%)	230 (38.5%)	117 (37.4%)	81 (36.0%)
potassium chloride, N (%)	1141 (32.1%)	661 (33.8%)	480 (30.0%)	219 (36.7%)	116 (36.9%)	75 (33.3%)
sodium chloride, N (%)	1134 (31.9%)	652 (33.4%)	482 (30.1%)	214 (35.8%)	112 (35.7%)	73 (32.4%)

Population characteristics includes patient demographic data as well as the top 4 most common Phenotype codes and 5 most common medication groups, stratified by overall, male, and female and are presented for the overall cohort and at 52 weeks.

Abbreviations: BMI, body mass index

Note: Data are N (%) for binary variables, and Mean (Standard Deviation) for continuous.

Table 2.

Weight loss of trackable individuals taking semaglutide

	Mean weight loss (kg)	Mean % weight loss	Lost >= 10% Weight	Lost >= 5% Weight
Overall				
Males, N=1,600	4.35 (3.96, 4.72)	3.66% (3.35, 3.97)	12.75% (11.12, 14.38)	36.44% (34.08, 38.80)
Females, N=1,955	5.09 (4.74, 5.45)	5.08% (4.74, 5.41)	22.46% (20.61, 24.30)	46.39% (44.18, 48.60)
Overall, N=3,555	4.76 (4.50, 5.02)	4.44% (4.21, 4.67)	18.09% (16.82, 19.35)	41.91% (40.29, 43.53)
12 weeks from semaglutide exposure				
Males, N=416	2.88 (2.48, 3.29)	2.46% (2.11, 2.80)	2.56% (1.13, 4.00)	21.79% (18.05, 25.54)
Females, N=603	3.65 (3.31, 3.99)	3.57% (3.24, 3.91)	6.90% (4.78, 9.01)	34.48% (30.51, 38.45)
Overall, N=1,019	3.30 (3.03, 3.56)	3.06% (2.82, 3.30)	4.91% (3.58, 6.23)	28.65% (25.88, 31.43)
52 weeks from semaglutide exposure				
Diabetes, N=450	4.19 (3.49, 4.88)	3.86% (3.26, 4.47)	14.00% (10.79, 17.21)	36.44% (32.00, 40.89)
Without Diabetes, N=89	7.94 (5.99, 9.88)	7.44% (5.73, 9.15)	33.71% (23.89, 43.53)	58.43% (48.19%, 68.67%)
Males, N=225	4.39 (3.42, 5.36)	3.76% (2.99, 4.52)	13.33% (8.89, 17.78)	33.33% (27.17, 39.49)
Females, N=314	5.10 (4.18, 6.03)	4.95% (4.10, 5.80)	20.06% (15.63, 24.49)	44.90% (39.40, 50.41)
Overall,N=539	4.80 (4.13, 5.47)	4.45% (3.87, 5.04)	17.25% (14.06, 20.44)	40.07% (35.94, 44.21)

Metrics describing overall weight loss and proportion of individuals meeting significant weight loss thresholds for the overall cohort and cohorts observed at endpoints 12 weeks and 52 weeks after beginning taking semaglutide. The Diabetes group includes individuals diagnosed with Type 1 or Type 2 diabetes.

Note: Data is in the form Mean (0.25%, 97.5%). Data includes the average weight loss as a percent lost of initial weight, while the bottom two rows are the proportion of patients in the entire cohort that achieve   5% and   10% weight loss.

Table 3.

Adjusted Odds Ratios of SHAP Important Features for Achieving 10% Weight Loss

RxNorm Ingredient Group, Phecode, or Demographic Factor	10% Weight Loss Odds Ratio (OR)	Number of Patients
dulaglutide	0.26 [0.09, 0.75]	92 (17.1%)
Diabetes mellitus	0.44 [0.22, 0.87]	450 (83.5%)
Age, Mean(SD)	0.98 [0.96, 0.99]	55.8 (12.1)
Disorders of the adrenal glands	25.45 [6.03, 107.45]	14 (2.6%%)
metoprolol	0.43 [0.18, 1.02]	110 (20.4%)
Initial weight (kg), Mean(SD)	1.00 [0.99, 1.00]	107.2 (25.6)
Disorders of the retina	0.23 [0.07, 0.73]	63 (11.7%%)
linaclotide	14.95 [2.54, 87.97]	10 (1.9%)
Elevated blood glucose level	2.32 [1.28, 4.20]	118 (21.9%)
Signs and symptoms involving emotional state	2.41 [0.90, 6.48]	25 (4.6%)
Other disorders of eye	25.50 [5.08, 128.12]	9 (1.7%)
Joint symptoms	0.62 [0.35, 1.12]	231 (42.9%)
metformin	0.55 [0.31, 0.97]	367 (68.1%)
Malnutrition and underweight	5.02 [1.59, 15.87]	20 (3.7%)
Lesions of mouth	8.38 [1.88, 37.37]	10 (1.9%)
meloxicam	0.34 [0.11, 1.08]	56 (10.4%)
codeine	2.97 [1.31, 6.73]	50 (9.3%)
Vitamin deficiencies	1.12 [0.62, 2.02]	206 (38.2%)
cough	0.86 [0.42, 1.73]	116 (21.5%)
atorvastatin	0.83 [0.46, 1.51]	245 (45.5%)

The adjusted odds ratios from a Logistic Regression model for the most important features obtained from SHAP analysis of the Catboost Classifier model.