Are Adverse Events of Semaglutide Different for Weight Control and Type II Diabetes?

Li Li
December 15, 2023
BST 260 Introduction to Data Science
SM 60 Biostatistics

1. Introduction

Semaglutide products were first approved as *Ozempic* by U.S. Food and Drug Administration (FDA) in 2017 to lower blood sugar levels in adults with type II diabetes mellitus (FDA, 2023). Semaglutide has a remarkable side effect, though: weight loss. One pivotal study in 2017 showed that the average losses for patients receiving one milligram a week of semaglutide were nearly 10 pounds over 30 weeks (Sorli et al., 2017). The side effects did not make people hesitate to use semaglutide products for treatment. Instead, not only people who want to lose weight without diabetes were attracted, but also Ozempic makers turned into a selling point. Eventually, accompanied by clinical trials, semaglutide received FDA approval in June 2021 for people with obesity or overweight, branded as *Wegovy* (FDA, 2021). Besides brands' advertisements, celebrities and social media widely shared weight loss experiences using semaglutide (New York Post, 2022). All of these accelerated forged prescriptions for people longing to lose weight with neither diabetes nor obesity, resulting in Wegovy and Ozempic shortage. With such promotion of semaglutide, the risks of misuse should be paid more attention to.

In recent years, the number of semaglutide prescriptions has increased substantially (Figure 1). Among all the patients prescribed semaglutide, while the share of patients using it for diabetes is decreasing, that for weight control is increasing (Figure 2). Current studies of semaglutide misuse focus on signals and symptoms at a biological level (Chiappini et al., 2023). There were no explicit comparison of adverse effect between the two reasons of use, nor adverse event count analysis controlled by the number of prescription. To better understand the potential difference between the adverse effect of semaglutide use for diabetes and for weight loss, this study tries to compare the adverse event rate adjusted for the number of prescription from 2020 to 2022, and compare how the adverse event characteristics were different in terms of age, sex, product, outcome, and seriousness from 2018 to 2022.

Three sources of data were used in this analysis. The FDA Adverse Event Reporting System (FAERS) database for semaglutide contains 19975 events and 22 covariates from 2017 to 2023, including reason for use, product name, age, sex, event date, etc. (FAERS, 2023). The ClinCalc DrugStats data for semaglutide contains the yearly number of prescriptions and patients from 2018 to 2020 (ClinCalc, 2023). The Trilliant Health page contains the percentage of patients prescribed with semaglutide with a history of diabetes or an eating disorder from Q1 2020 to Q3 2023 (Trilliant Health, 2023).

2. Methods

2.1. Data Wrangling and Web Scrapping

All statistical analyses were conducted using R (version 4.2.0) and RStudio (version 2023.02.0).

The columns of data from the FAERS database (FAERS, 2023) were cleaned and categorized. The number and units in the columns of `Weight` and `Age` were extracted respectively and converted to integer numbers in unified units of kilograms and years. The mean age is 62 years, and the mean weight is 96.94 kg. The `Date` was organized using `lubridate` package. The text strings in `Reason of Use`, `Outcomes`, and `Suspect Product` were split, extracted by keyword, and categorized into desired levels. `Reason of Use` was categorized into diabetes (53.9%), weight loss (9.3%), and others (36.9). `Outcomes` were categorized into died, hospitalized, and others. `Suspect Product` was categorized into the current most popular three products: Ozempic (injection for diabetes, 73.9%), Rybelsus (tablet for diabetes, 10.8%), and Wegovy (injection for obesity, 7.8%). The data was filtered to include only data from 2018 to 2022 because of delayed reporting in 2023 and limited sample size in 2017 and 2023.

The yearly overall prescription number was scrapped from ClinCalc (ClinCalc, 2023) using `rvest` package. The data only consists of year 2018 to 2020 and the overall number of patients prescribed with semaglutide. The quarterly prescription number change was scrapped from Trilliant Health (Trilliant Health, 2023). The data only consists of time Q1 2020 to Q3 2022. Moreover, the data does not have an overall number of patients, only percentage changes compared to Q1 2020 for overall patient number and share (i.e., percentage) of patients for diabetes and weight control. Because of the relatively smooth trends (Figure 2), the Q4 2022 data was imputed using generalized additive models (GAM) ($share \sim f(time)$) to enable the calculation of the yearly trend. The specific number of patients for year 2021 and 2022 was extrapolated from the ClinCalc data based on percent change data from Trilliant Health, as the 2020 data is overlapped. The trend of the overall adverse event rate was then compared from 2018 to 2022. The trend of adverse event rate was compared for the two reasons of use for 2020 to 2022 because of the limited share of patient data.

2.2. Logistic Regression

In the outcome variable `Reason of Use`, 53.9% is diabetes, 9.3% is weight loss, and 36.9% is others. The majority of others is simply not specified and thus not indicative. Others also include reasons like dyspepsia (indigestion), glucose tolerance impaired, glycosylated hemoglobin increased, etc., which are diet, glucose, and insulin-related, but can not be clearly classified without more details provided. The Others in `Reason of Use` was therefore treated as missing data. The missingness in covariates is summarized in Table 1. The missingness in `Reason of Use`, i.e., if the reason is others, was modeled with logistic regression with covariates with less than 25% missingness: $logit(p_{missing}) = \beta_0 + \beta_1 product + \beta_2 sex + \beta_3 year + \beta_4 serious + \beta_5 country + \beta_6 outcomes + \beta_7 age$

The fitted model suggested that the missingness in `Reason of Use` has a statistically significant association with `Year` and `Product`. Therefore, the missingness is not missing completely at random (MCAR). However, there is no empirical evidence supporting that the missingness is related to any specific reason of use. Therefore, we assumed that the missingness is missing at random (MAR), conditioning on `Year` and `Product`. The following modeling then used complete case analysis. The `Weight` covariate was dropped because of too much missingness (77.2%).

A logistic regression model was used to analyze the relationship between the binary outcome `Reason of Use` and several characteristics of the adverse event, including `Sex`, `Age`, `Serious`, and `Outcomes`, adjusted for `Product` and `Year` as suggested by the MAR analysis above. The model analyzed is shown below as Model 1. `Year` was modeled flexibly to be better controlled. The interpretability is unimportant compared to other covariates as it is not the main interest in this research question. The coefficients were analyzed, and the model performance was analyzed using ROC AUC.

Model 1:
$logit(p_{weightLoss}) = \beta_0 + \beta_1 product + \beta_2 sex + \beta_3 f(year) + \beta_4 serious + \beta_5 outcomes + \beta_6 age$

3. Results

3.1 Adverse Event Rate Trend

After the data wrangling with the FAERS data, the trend of adverse event count from 2018 to 2022 is shown in Figure 3. The trend of the overall count was increasing and peaked in 2021. For the specific reason of use as weight control, the count was monotonically increasing through the five years, with the largest increase happening in 2021. This is not surprising because semaglutide was approved for weight control use in June 2021. For diabetes, the count was decreasing after 2019.

After web-scrapping and extrapolating the ClinCalc data using the Trilliant health data and imputation of Q4 2022 with GAM, the trend of the number of patients prescribed with semaglutide from 2018 to 2022 is shown in Figure 4. The trend has increased dramatically over the five years. Immediately, it is noticeable that the increasing trend of adverse event numbers is not as sharp as that of prescription numbers over the years.

The overall rate of adverse events from 2018 to 2022 is shown in Figure 5. Slightly unexpectedly, the rate was decreasing monotonically. This might be because as the main share of patients prescribed with semaglutide remains to be for diabetes (~70%), the adverse event rate for those patients decreases due to possible reasons like more standardized diagnosis and prescriptions and more complete regulations for semaglutide products, since Ozempic was approved just after 2017. Changes in other shares of patients may not have an impact influential enough on the overall rate change compared to the relatively large absolute number of patients using semaglutide for diabetes.

The rate of adverse events for diabetes and weight control respectively from Q1 2020 to Q4 2022 is shown in Figure 6. In general, the adverse event rate for weight control was higher than that for diabetes. While the rate of diabetes decreased through the three years, the rate for weight control marked its highest increase from Q2 to Q3 2021, which was exactly when semaglutide was approved for weight control in June 2021. It is worth noticing that after Q3 2021, the trend started to decrease, meaning that the subsequent promotion or celebrity effect on social media was not accompanied by an increase in adverse event rate. Similar to use for diabetes, the adverse event rate usually had its peak when the drug was initially approved but was not greatly affected by other factors (increased prescription and promotion on social media). Despite this, the generally higher adverse event rate for weight control still raises concerns about semaglutide products for the specific use of weight control.

3.2 Adverse Event Characteristics Modeled by Logistic Regression

Model 1:

$$logit(p_{weightLoss}) = \beta_0 + \beta_1 product + \beta_2 sex + \beta_3 f(year) + \beta_4 serious + \beta_5 outcomes + \beta_6 age$$

Model 1 was fitted to analyze the characteristic difference for adverse events between the two reasons of use (diabetes and weight control) from 2018 to 2022. The fitted results of Model 1 are shown in Table 2. For both reasons of use, the coefficients for `Serious` and `Outcomes` (died, hospitalized, and others) are not significant, indicating that there is not enough evidence to show the severity and outcome of adverse events were different for the two reasons of use.

Product is one of the covariates with significant coefficients. The most obvious one is that patients using Wegovy had an odds of using it for weight control 126.3 (95% CI [125.7, 127.0]) times the odds for patients using Ozempic, on average according to these data, since Wegovy is intended for weight control. Patients using Rybelsus had an odds of using it for weight control 0.197 (95% CI [0, 0.714]) times the odds for patients using Ozempic, suggesting that among the adverse events, while patients still used Ozempic (intended for diabetes) for weight control, patients using Rybelsus (intended for diabetes) seemed to be more likely to use it just for diabetes. Last but not least, patients using semaglutide products other than those three had an odds of using it for weight control 6.195 (95% CI [5.452, 6.938]) times the odds for patients using Ozempic, on average, according to these data.

Among other significant covariates, females had higher odds of using semaglutide for weight control compared to males. The result is highly significant, and the odds ratio is far from 1. Additionally, the age distribution of the adverse events is approximately normal, with a considerably high mean of 62 years. Among them, younger people had higher odds of using semaglutide for weight control, on average according to these data.

The ROC curve of Model 1 is shown in Figure 6. It reaches an AUC of 0.8979, which indicates high goodness-of-fit with respect to the model performance.

4. Conclusion

The study uses three data sources (FAERS, CLinCalc, and Trilliant Health) to analyze the trend of adverse event rates among prescriptions and the characteristics of adverse events, with a comparison between use for diabetes and use for weight control. The adverse event rate for weight control was higher than that for diabetes from 2020 to 2022. While the rate decreased monotonically for diabetes, that for weight control increased most from Q2 to Q3 in 2021, exactly when semaglutide was first approved for weight control treatment. The logistic regression model suggested that among the adverse events from 2018 to 2022, younger females using Wegovy and products other than Ozempic and Rybelsus tend to be more likely to use semaglutide for weight control. The ROC AUC of the model was 0.8979. No part of the analysis suggests a causal relationship.

One of the limitations of the study is that in FAERS, how long reporting is delayed needs to be clarified. This may affect the accuracy of adverse event counts, especially in recent years. For example, the decrease in adverse event count in 2022 can be partly due to the delayed reporting problem, but there is no evidence. The possible bias brought by the self-reporting system of FAERS is also unclear and can be a concern. Because of the nature of the three different sources, the data collection methods might be slightly inconsistent among the three datasets used and therefore affecting the accuracy of the specific numbers in the final results of the adverse event rate.

Future directions of the study involve collecting the number of prescriptions with details of the reason of use by a more unified method from 2018 to 2022 and adding those details to this analysis. The delayed reporting problem can also be more systematically analyzed to predict how long an adverse event is expected to be delayed.

5. References

Chiappini, S., Vickers-Smith, R., Harris, D. R., Pelletier, G., Corkery, J., Guirguis, A., Martinotti, G., Sensi, S. L., & Schifano, F. (2023). Is There a Risk for Semaglutide Misuse? Focus on the Food and Drug Administration's FDA Adverse Events Reporting System (FAERS) Pharmacovigilance Dataset. *Pharmaceuticals, 16*(7), 994–994. https://doi.org/10.3390/ph16070994

FDA Approves New Drug Treatment for Chronic Weight Management, First Since 2014. (2021). FDA. https://www.fda.gov/news-events/press-announcements/fda-approves-new-drug-treatment-chronic-weight-management-first-2014

Hollywoods New Secret tp Losing Weight is a Diabetes Injection. (2022). New York Post. https://nypost.com/2022/10/11/hollywoods-new-secret-to-losing-weight-is-a-diabetes-injection/

Medications Containing Semaglutide Marketed for Type 2 Diabetes or Weight Loss. (2023). FDA. https://www.fda.gov/drugs/postmarket-drug-safety-information-patients-and-providers/medications-containing-semaglutide-marketed-type-2-diabetes-or-weight-loss

Sorli, C., Harashima, S., Tsoukas, G. M., Unger, J., Karsbøl, J. D., Hansen, T., & Bain, S. C. (2017). Efficacy and safety of once-weekly semaglutide monotherapy versus placebo in patients with type 2 diabetes (SUSTAIN 1): a double-blind, randomised, placebo-controlled, parallel-group, multinational, multicentre phase 3a trial. *The Lancet Diabetes & Endocrinology, 5*(4), 251–260. https://doi.org/10.1016/s2213-8587(17)30013-x

Data

FDA Adverse Event Reporting System (FAERS) Public Dashboard. (2023). FDA. https://fis.fda.gov/sense/app/95239e26-e0be-42d9-a960-9a5f7f1c25ee/sheet/45beeb74-30ab-46be-8267-5756582633b4/state/analysis

Patients Prescribed Drugs Like Ozempic and Mounjaro Have Increased Over 300%. (2023). Trilliant Health. https://www.trillianthealth.com/insights/the-compass/patients-prescribed-drugs-like-ozempic-and-mounjaro-have-increased-over-300-percent

Semaglutide Drug Usage Statistics. (2023). Clincalc.
https://clincalc.com/DrugStats/Drugs/Semaglutide

6. Appendix

6.1 Code

```r
library(tidyverse)
library(readxl)
library(lubridate)
library(rvest)
library(vctrs)
library(gam)
library(GGally)
library(caret)
library(gridExtra)
library(grid)
library(pROC)
```

```r
# FAERS
dat = read_xlsx("FAERS_Semaglutide.xlsx")
```

```r
# data cleaning
# clean reason for use
ifdiab = function(s) {grepl("Diabetes", s, fixed=TRUE)}

ifwl = function(s) {grepl("Weight", s, fixed=TRUE) |
    grepl("weight", s, fixed=TRUE) |
    grepl("Obesity", s, fixed=TRUE)}

# clean suspect product of use
whichprod = function(s) {
  out = rep(NA, length(s))
  for (i in 1:length(s)){
    if (grepl("-", s[i], fixed=TRUE)) {out[i] = NA}
    else if (grepl("Wegovy", s[i], fixed=TRUE)) {out[i] = "Wegovy"}
    else if (grepl("Ozempic", s[i], fixed=TRUE)) {out[i] = "Ozempic"}
    else if (grepl("Rybelsus", s[i], fixed=TRUE)) {out[i] = "Rybelsus"}
    else {out[i] = "Others"}
  }
  out
}
```

```r
# clean patient age with various units in a string
whichage = function(s) { # in years
  out = rep(NA, length(s))
  for (i in 1:length(s)){
    if (grepl("Not Specified", s[i], fixed=TRUE)) {out[i] = NA}
    else if (grepl("YR", s[i], fixed=TRUE)) {
      out[i] = as.numeric(gsub("\\D", "", s[i]))}
    else if (grepl("MTH", s[i], fixed=TRUE)) {
      out[i] = round(as.numeric(gsub("\\D", "", s[i])) / 12)}
    else if (grepl("DEC", s[i], fixed=TRUE)) {
      out[i] = as.numeric(gsub("\\D", "", s[i])) * 10}
    else if (grepl("DAY", s[i], fixed=TRUE)) {
      out[i] = round(as.numeric(gsub("\\D", "", s[i])) / 365)}
    else if (grepl("WEEK", s[i], fixed=TRUE)) {
      out[i] = round(as.numeric(gsub("\\D", "", s[i])) / 52)}
    else {out[i] = 999} # for debug use
  }
  as.numeric(out)
}

# clean patient weight with various units in a string
whichwt = function(s) { # in kg
  out = rep(NA, length(s))
  for (i in 1:length(s)){
    if (grepl("Not Specified", s[i], fixed=TRUE)) {out[i] = NA}
    else if (grepl("KG", s[i], fixed=TRUE)) {
      out[i] = as.numeric(str_extract(s[i], "\\d+\\.*\\d*"))}
    else if (grepl("LB", s[i], fixed=TRUE)) {
      out[i] = as.numeric(str_extract(s[i], "\\d+\\.*\\d*")) * 0.45}
    else {out[i] = 999} # for debug use
  }
  as.numeric(out)
}

# clean outcome to levels: died, hospitalized, others

whichoutcome = function(s) {
  out = rep(NA, length(s))
  for (i in 1:length(s)){
    if (grepl("Died", s[i], fixed=TRUE)) {out[i] = "Died"}
    else if (grepl("Hospitalized", s[i], fixed=TRUE)) {
      out[i] = "Hospitalized"}
    else {out[i] = "Others"}
  }
  out
}

# cleaned dataset:
dat1 = dat %>% mutate(reason = ifelse(ifdiab(`Reason for Use`),
                                      "Diabetes",
                                      ifelse(ifwl(`Reason for Use`),
                                             "Weight", "Others"))) %>%
  rename("product" = "Suspect Product Names",
         "date" = "Event Date",
         "age" = "Patient Age",
         "weight" = "Patient Weight",
         "country" = "Country where Event occurred") %>%
  mutate(country = ifelse(country == "Not Specified", NA, country),
         product = whichprod(product),
         age = whichage(age),
         weight = whichwt(weight),
         year = as.numeric(str_extract(date, "20\\d+")), # clean date
         month = match(str_extract(date, "[[:alpha:]]+"),
                       toupper(month.abb)),
         date = dmy(date),
         Sex = ifelse(Sex == "Not Specified", NA, Sex),
         Outcomes = whichoutcome(Outcomes)) %>%
  mutate(country = as.factor(country),
         reason = factor(reason, levels = c("Others",
                                            "Weight", "Diabetes")),
         product = factor(product, levels = c("Ozempic", "Wegovy",
                                              "Rybelsus", "Others")),
         Serious = as.factor(Serious),
         Outcomes = factor(Outcomes, levels = c("Others", "Hospitalized",
                                                "Died")),
         Sex = as.factor(Sex)) %>%
  select(product, reason, Serious, Outcomes, Sex, date, year,
         month, age, weight, country) %>%
  filter(year >= 2018 & year <= 2022) %>%
  mutate(quarter = ceiling(month / 3))
# because of sample size: study 2018-2022
# for debug use:
# dat[which(dat1$weight == 999),]
```

```
p.advTrend = dat1 %>%
  filter(country == "US") %>%
  ggplot(aes(x = year, fill = reason)) +
  geom_bar() +
  scale_fill_manual(values = c("grey", "aquamarine2", "aquamarine4")) +
  theme_bw() +
  labs(title = "Number of adverse event of Semaglutide over years",
       subtitle = "Stratified by reason of use")
```

```
# web scrapping of prescription 2018-2020
# Clinicalc
clincalc <- read_html("https://clincalc.com/DrugStats/Drugs/Semaglutide")
```

```
clincalc.fig1 = html_nodes(clincalc, "script")[5] %>% html_text

presc = str_extract_all(clincalc.fig1, "20\\d{2}.*?]")[[1]] %>%
  str_extract_all("\\d*") %>% as.data.frame
colnames(presc) = c(1, 2, 3)
presc = t(presc)[, c(1, 4, 6)] %>% as.data.frame
colnames(presc) = c("year", "n_prescriptions_yr", "n_patients_yr")
presc = mutate(presc, year = as.numeric(year),
               n_prescriptions_yr = as.numeric(n_prescriptions_yr),
               n_patients_yr = as.numeric(n_patients_yr))
```

```
# web scrapping of patients 2020-2022
# Trilliant

# hard coded because sadly found the data on web is png
trilliant = data.frame(year = c(2020, 2020, 2020, 2020,
                                2021, 2021, 2021, 2021,
                                2022, 2022, 2022, 2022),
                       #percent change compared to Q1 2020
                       change = c(0, 0.13, 0.25, 0.36,
                                  0.53, 0.79, 1.17, 1.53,
                                  1.94, 2.52, 3.55, NA),
                       quarter = c(1, 2, 3, 4,
                                   1, 2, 3, 4,
                                   1, 2, 3, 4),
                       share_diab = c(0.767, 0.772, 0.774, 0.775,
                                      0.773, 0.761, 0.705, 0.666,
                                      0.634, 0.615, 0.570, NA),
                       share_wt = c(0.0044, 0.0046, 0.0049, 0.0051,
                                    0.0058, 0.0070, 0.0079, 0.0085,
                                    0.0088, 0.0092, 0.0093, NA))

# impute 2022 Q4 change based on GAM regression
# for further investigation.
timestamp = 1:11
mod.c = gam(trilliant$change[1:11] ~ s(timestamp),
           data = trilliant)
trilliant[12, 2] = predict(mod.c, newdata = data.frame(timestamp = 12))
mod.diab = gam(trilliant$share_diab[1:11] ~ s(timestamp),
           data = trilliant)
trilliant[12, 4] = predict(mod.diab, newdata = data.frame(timestamp = 12))
mod.wt = gam(trilliant$share_wt[1:11] ~ s(timestamp),
           data = trilliant)
trilliant[12, 5] = predict(mod.wt, newdata = data.frame(timestamp = 12))

# manipulation: convert change into n_patients based on clincalc info
presc2 = trilliant %>% mutate(change = change + 1) %>%
  group_by(year) %>%
  summarize(change = sum(change))

presc.quarter = trilliant %>% mutate(change = change + 1) %>%
  mutate(n_patients_diab_q = round(share_diab * change *
           as.numeric(presc[3, 3])/as.numeric(presc2[1, 2])),
         n_patients_wt_q = round(share_wt * change *
           as.numeric(presc[3, 3])/as.numeric(presc2[1, 2]))) %>%
  select(year, quarter, n_patients_diab_q, n_patients_wt_q)

presc2 = presc2 %>%
  mutate(n_patients_yr = change *
           as.numeric(presc[3, 3])/as.numeric(presc2[1, 2])) %>%
  select(-change) %>%
  rbind(presc[-3, -2])
dat2 = dat1 %>%
  left_join(presc2, by = "year") %>%
  left_join(presc.quarter, by = c("year", "quarter"))
```

```
# Figure 2
years <- c("2020", "2021", "2022")
quarters <- c("Q1", "Q2", "Q3", "Q4")
labels <- c()
for (year in years) {
    for (quarter in quarters) {
        labels <- c(labels, paste0(year, " ", quarter))
    }
}

p.A = trilliant %>% mutate(timestamp = 1:12) %>%
  filter(timestamp != 12) %>%
  ggplot(aes(x = timestamp, y = share_diab)) +
  geom_bar(stat = "identity", fill = "aquamarine4") +
  theme_bw() +
  labs(x = "time", y = "share",
       subtitle = "For diabetes") +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_x_continuous(breaks = 1:11, labels = labels[1:11]) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5))

p.B = trilliant %>% mutate(timestamp = 1:12) %>%
  filter(timestamp != 12) %>%
  ggplot(aes(x = timestamp, y = share_wt)) +
  geom_bar(stat = "identity", fill = "aquamarine2") +
  theme_bw() +
  labs(x = "time", y = "share",
       subtitle = "For weight control") +
  scale_y_continuous(labels = scales::percent_format(),
                     limits = c(0, 0.02)) +
  scale_x_continuous(breaks = 1:11, labels = labels[1:11]) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5))
```

```
# Figure 1
p1 = trilliant %>% mutate(timestamp = 1:12) %>%
  filter(timestamp != 12) %>%
  ggplot(aes(x = timestamp, y = change)) +
  geom_line(stat = "identity") +
```

```r
  geom_point() +
  theme_bw() +
  labs(x = "time", y = "percent change",
       title = "Percent change of number of patients prescirbed with Semaglutide",
       subtitle = "Compared to Q1 2020") +
  scale_x_continuous(breaks = 1:11, labels = labels[1:11]) +
  theme(axis.text.x = element_text(angle = 45, vjust = 0.5))
```

```r
p.prescTrend = dat2 %>% group_by(year) %>%
  filter(country == "US") %>%
  summarize(n_patients_yr = unique(n_patients_yr)) %>%
  ggplot(aes(x = year, y = n_patients_yr)) +
  geom_bar(stat='identity') +
  labs(y = "count",
       title = "Number of patients prescribed with Semaglutide over years",
       subtitle = "") +
  theme_bw()
```

```r
p.rateAll = dat2 %>% group_by(year) %>%
  filter(country == "US") %>%
  summarize(n_patients_yr = unique(n_patients_yr),
            n_adv = n()) %>%
  ggplot(aes(x = year, y = n_adv/n_patients_yr)) +
  geom_line(stat='identity') +
  geom_point() +
  labs(y = "rate",
       title = "Rate of adverse event of Semaglutide over years",
       subtitle = "Per patient prescribed") +
  theme_bw()
```

```r
# now shift our focus to reasons of use
# because of limited data, we can only analyze 2020 - 2022 data

years <- c("2020", "2021", "2022")
quarters <- c("Q1", "Q2", "Q3", "Q4")
labels <- c()
for (year in years) {
    for (quarter in quarters) {
        labels <- c(labels, paste0(year, " ", quarter))
    }
}

# new dat dimension after getting rid of NA: 3996 x 16
p.rateDiv = dat2 %>% filter(year >= 2020) %>%
  filter(country == "US") %>%
  mutate(timestamp = (year - 2020)*4 + quarter) %>%
  filter(reason != "Others") %>%
  filter(!is.na(timestamp)) %>%
  group_by(timestamp, year, quarter, reason, n_patients_diab_q,
           n_patients_wt_q) %>%
  summarize(count = n(),
            .groups = "drop") %>%
  mutate(rate = ifelse(reason == "Diabetes", count/n_patients_diab_q,
                       count/n_patients_wt_q)) %>%
  ggplot(aes(x = timestamp, y = rate, group = reason, color = reason)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = 1:12,
                     labels = labels) +
  scale_color_manual(values = c("aquamarine2", "aquamarine4")) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, vjust = 0.5)) +
  labs(x = "time",
       title ="Rate of adverse event of Semaglutide over years",
       subtitle = "Per patient prescribed, stratified by reason of use")
```

```
missing = dat2 %>%
  mutate(isOthers = ifelse(reason == "Others", 1, 0)) %>%
  select(product, reason, Serious, Outcomes, Sex, year,
         age, weight, country, isOthers)

tbl_missing = missing %>% apply(2, FUN = function(col) {
  mean(is.na(col))
})

# select columns with at least 90% data not missing
# dropped Outcomes because too many levels
missing.mod = glm(isOthers ~ product + Serious +
                    Sex + year + country + Outcomes + age,
                  family = binomial(), data = missing)
#summary(missing.mod)
```

```
# final data set for regression analysis
dat3 = missing %>%
  select(product, reason, Serious, Outcomes,
                          Sex, year, age) %>%
  filter(reason != "Others") %>%
  filter(!is.na(product) & !is.na(Sex) & !is.na(age)) %>%
  mutate(reason = ifelse(reason == "Diabetes", 0, 1))
```

```
# reason: 1-weight, 0-diabetes
mod1 = glm(as.factor(reason) ~ product + Serious + Outcomes + Sex
           + s(year) + age,
           family = binomial(), data = dat3)

# summary(mod1)

p <- predict(mod1, type = "response")
pred <- as.factor(round(p))

dat3.pred = dat3 %>% mutate(logit = log(p/(1-p)),
                            pred = pred)

roccurve <- roc(dat3$reason ~ p)

# auc(roccurve)
# Area under the curve: 0.8979
```

```
coef = exp(summary(mod1)$coefficients[2:10, 1])
sd = summary(mod1)$coefficients[2:10, 2]
cilo = round(coef - 1.96*sd, 3)
cilo[cilo < 0] = 0.000
cihi = round(coef + 1.96*sd, 3)
p = round(summary(mod1)$coefficients[2:10, 4], 3)

coef = round(coef, 3)

mod1.out = data.frame(expb = coef,
                      ci = paste0("[", cilo, ", ", cihi, "]"),
                      p = ifelse(p<0.001, "< 0.001", as.character(p)))
colnames(mod1.out) = c("exp(coefficient)", "95% confidence interval", "p-value")
```

6.2 Figures and Tables

Table 1. Missingness in FAERS data.

variable	percent.missing
product	6.7%
reason	0%
serious	0%
outcomes	0%
sex	1.4%
year	0%
age	23.8%
weight	77.2%
country	3.2%
isothers	0%

Table 2. Exponential coefficient (i.e. odds ratio) of Model 1. Reference group for product is Ozempic, for outcomes is Others.

	exp(coefficient)	95% confidence interval	p-value
productWegovy	126.337	[125.678, 126.996]	< 0.001
productRybelsus	0.197	[0, 0.714]	< 0.001
productOthers	6.195	[5.452, 6.938]	< 0.001
SeriousSerious	0.898	[0.631, 1.165]	0.431
OutcomesHospitalized	0.960	[0.655, 1.265]	0.792
OutcomesDied	0.959	[0, 2.017]	0.938
SexMale	0.423	[0.163, 0.683]	< 0.001
s(year)	2.006	[1.907, 2.105]	< 0.001
age	0.939	[0.931, 0.948]	< 0.001

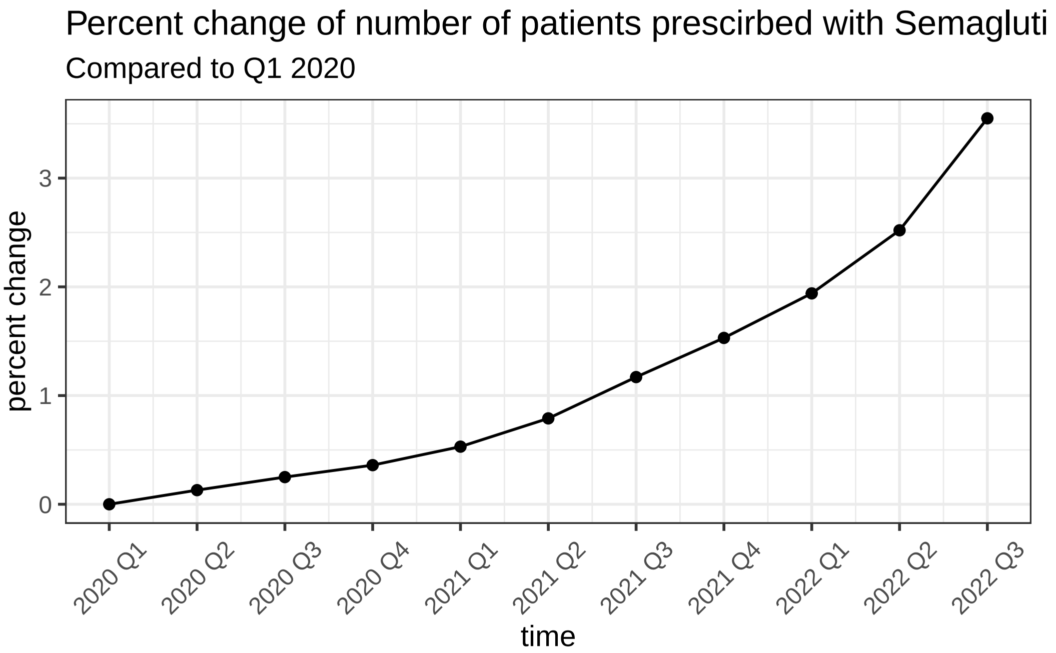

Figure 1. Percent change of number of patients prescribed with Semaglutide (compared to Q1, 2020). Source: Trilliant Health, 2023.

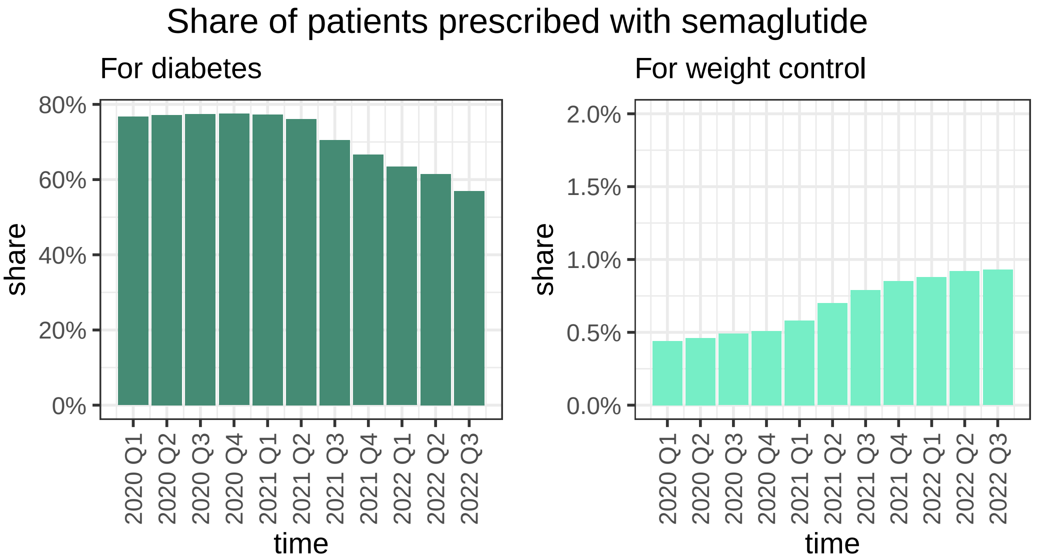

Figure 2. Share of patients prescribed with semaglutide comparison for diabetes and weight control. Source: Trilliant Health, 2023.

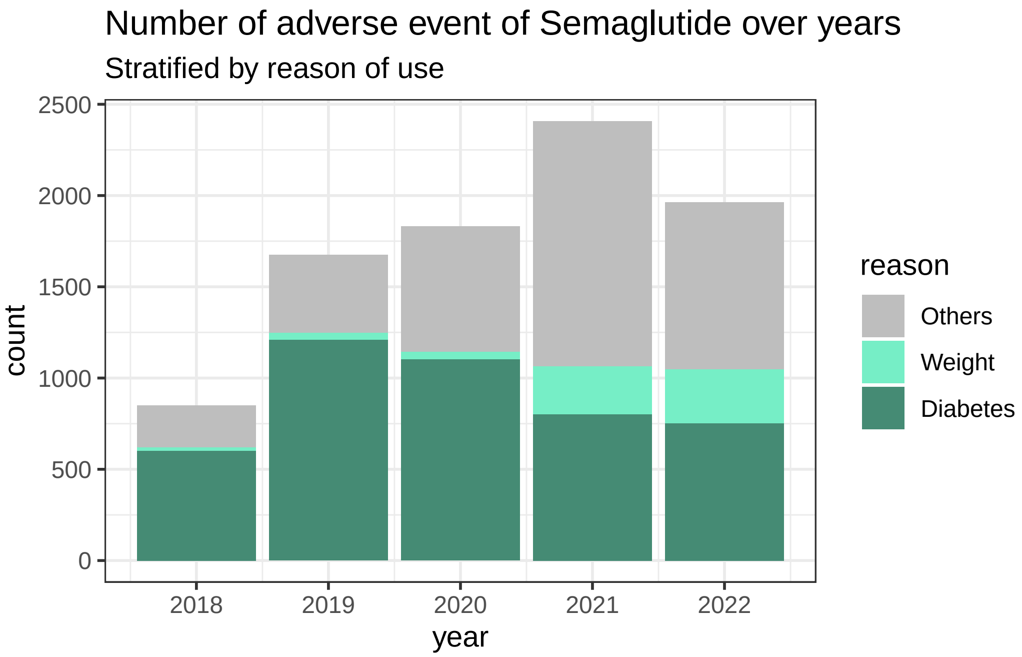

Figure 3. Number of adverse event of Semaglutide trend from 2018 to 2022 in the US, with stratified reason of use.

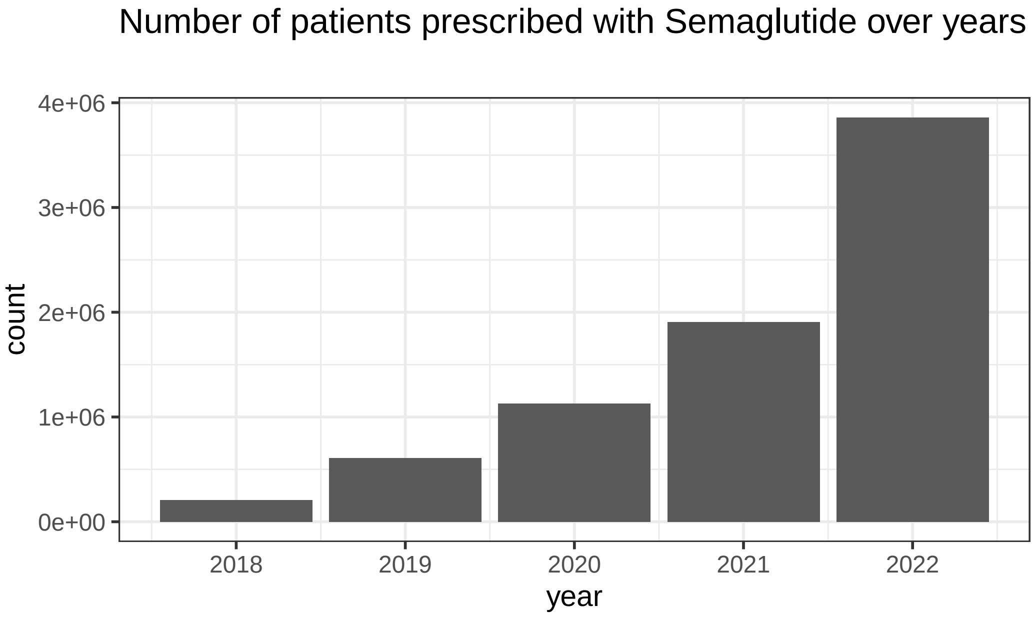

Figure 4. Number of patients prescribed with Semaglutide trend from 2018 to 2022 in the US.

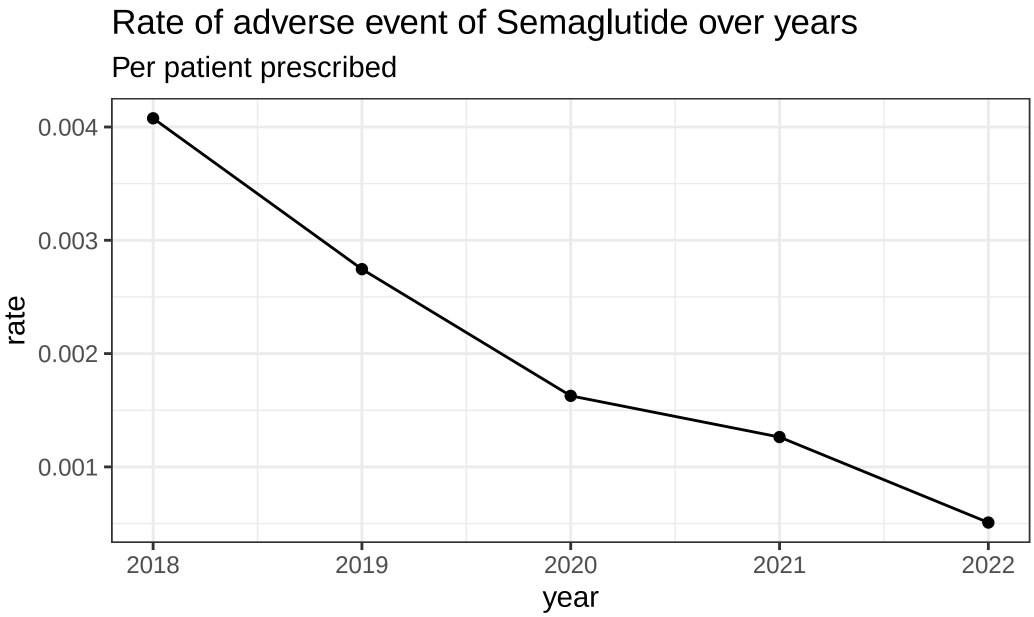

Figure 5. Rate of adverse event (count per patient prescribed with Semaglutide) trend from 2018 to 2022 in the US.

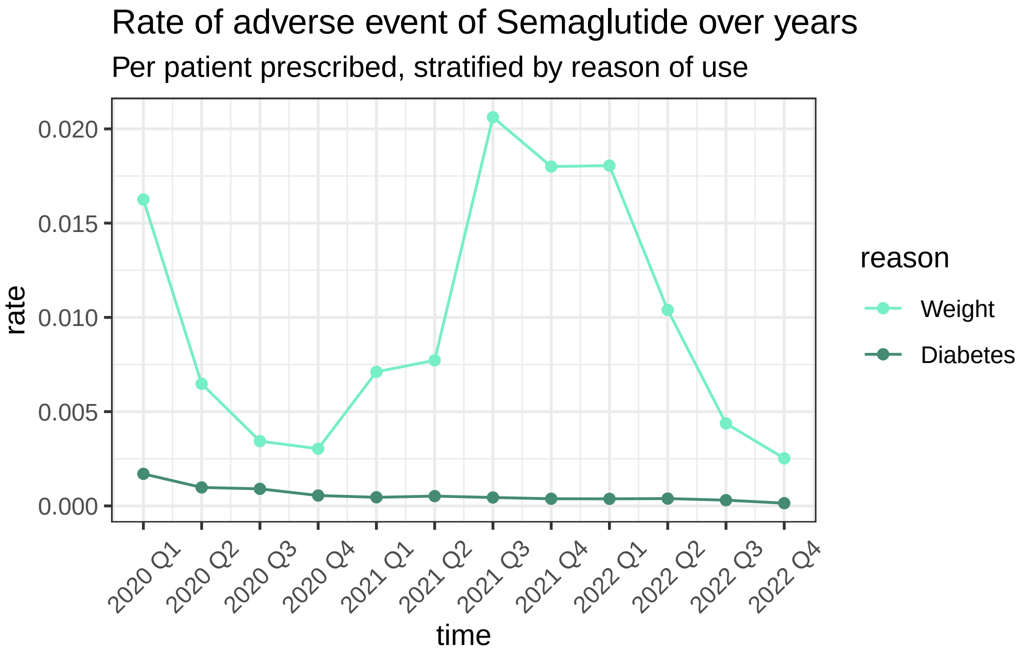

Figure 6. Comparison of rate of adverse event (count per patient prescribed with Semaglutide) for weight control and diabetes from Q1 2020 to Q4 2022 in the US.

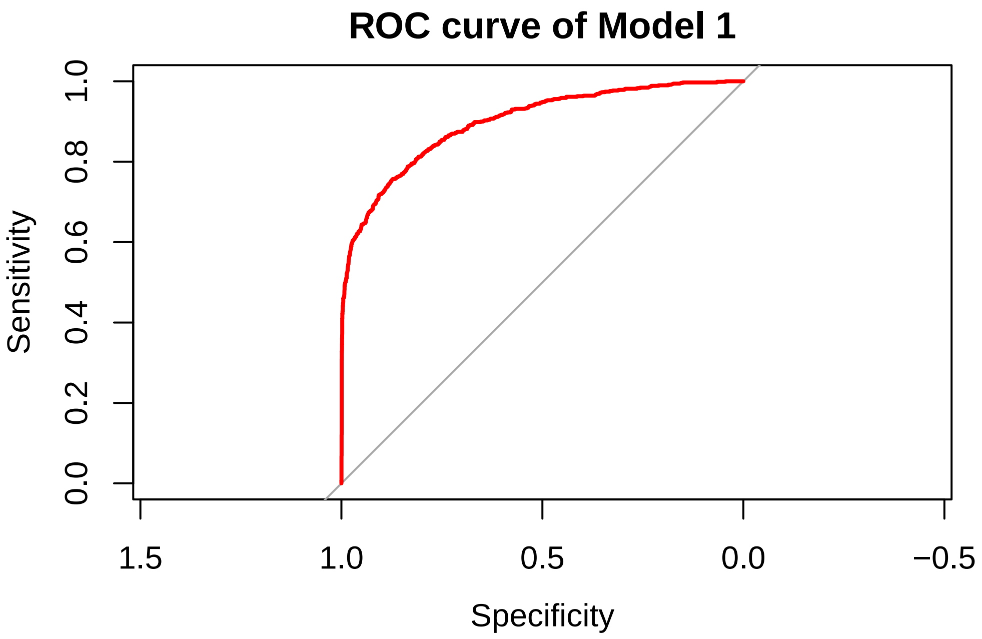

Figure 7. ROC curve of Model 1, AUC = 0.8979.