Assessing the Use of Prescription Drugs in Obese Respondents in the National Health and Nutrition Examination Survey

Laura A. Barrett, MS¹, Aiwen Xing, MS², Julia Sheffler, PhD³, Elizabeth Steidley¹, Terrence J. Adam, RPh, MD, PhD⁴, Rui Zhang, PhD⁴, Zhe He, PhD¹,³

¹School of Information, Florida State University, Tallahassee, FL, USA
²Department of Statistics, Florida State University, Tallahassee, FL, USA
³Department of Behavioral Sciences and Social Medicine, College of Medicine, Florida State University, Tallahassee, FL, USA
⁴Institute for Health Informatics and College of Pharmacy, University of Minnesota, Minneapolis, MN, USA

Corresponding Author:

Zhe He, PhD
School of Information
College of Communication and Information
Florida State University
142 Collegiate Loop
Tallahassee, Florida 32306-2100
zhe@fsu.edu
Phone: 001(850)644-5775

Competing Interests
None

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Introduction: Obesity is a common disease and a known risk factor for many other conditions such as hypertension, type 2 diabetes, and cancer. Treatment options for obesity include lifestyle changes, pharmacotherapy, and surgical interventions such as bariatric surgery. In this study, we examine the use of prescription drugs and dietary supplements by the individuals with obesity.

Methods: We conducted a cross-sectional analysis of the National Health and Nutrition Examination Survey (NHANES) data 2003-2014. We used multivariate logistic regression to analyze the correlations of demographics and obesity status with the use of prescription drugs and dietary supplement use. We also built machine learning models to classify prescription drug and dietary supplement use using demographic data and obesity status.

Results: Individuals with obesity are more likely to take cardiovascular agents (OR=1.265, 95% CI 1.222-1.311) and metabolic agents (OR=1.398, 95% CI 1.343-1.456) than individuals without obesity. The best performing model for classifying prescription drug use had the accuracy of 74.5% and the AUROC of 0.817.

Conclusions: This study can inform clinical practice and patient education of the use of prescription drugs and dietary supplements and their correlation with obesity.
Introduction

As a major health and economic crisis affecting the modern world, much progress has been made in identifying and developing strategies for preventing and treating obesity. Currently, treatment options include lifestyle changes, pharmacotherapy, and surgical interventions (e.g., intragastric balloons and bariatric surgery). In terms of pharmacotherapy, there are five approved prescription drugs (RXD) (orlistat, 1999; phentermine/topiramate, 2012; liraglutide, 2014; naltrexone/bupropion, 2014; and semaglutide, 2021) that can be prescribed for weight loss. All but orlistat, which reduces the absorption of fat, work by helping the individual to limit caloric intake. There are also four RXD that are similar to amphetamines that can be used short-term (phendimetrazine, diethylpropion, phentermine, and benzphetamine). There have been three other well-known RXD that were approved for use and then removed from the market. The first one is fenfluramine/phentermine (fen-phen) which was discontinued in 1997 because fenfluramine was shown to cause cardiac issues. The second one is sibutramine, which was withdrawn in 2010 due to an increased risk of stroke and myocardial infarction. The third one is lorcaserin, which was withdrawn in February 2020 after a clinical trial showed an increased occurrence of cancers. Due to the cost of pharmacotherapy and surgical interventions, as well as other reasons, dietary supplements (DS) are often used as a cost-sensitive and easily accessible, albeit less scientifically supported, alternative treatment of obesity.

Individuals with obesity face an increased risk of chronic diseases, namely depression, type 2 diabetes, cardiovascular disease, and many cancers including those of the colon, breast, kidney, and pancreas. Many of these conditions require pharmaceutical intervention as part of the treatment plan and individuals with obesity often use RXDs to manage these conditions. Overall, RXD use in the United States has increased. This increase is partly influenced by the
Development of new RXDs, the expansion of RXD coverage by insurance companies, and increased rates of chronic conditions such as obesity. The greatest increase in RXD use has been in those used for treating conditions found to be associated with obesity, specifically antihypertensives, antihyperlipidemic, antidiabetics, and antidepressants. Recent studies have researched obesity in relation to specific drugs or drug types. There have also been recent studies that examined various aspects of obesity such as childhood obesity, obesity and hours spent at work, exposure to certain pollutants or chemicals, trends in obesity, and obesity and waist circumference. However, there have not been any studies that look at overall RXD use in individuals with obesity. Being able to see this bird’s eye view of this relationship is important because understanding the patterns of RXD use among people with obesity, who often have other chronic conditions, can inform both clinical practice and research. This is challenging because there are cross-over issues between RXD, their side effect of weight gain, and their therapeutic effect on obesity and its comorbidities. For example, certain blood-glucose-lowering RXD and psychotropics may lead to unintended weight gain. In this project, we aim to gain an in-depth understanding of both the relationship between obesity and RXD use, as well as the correlations between specific RXD and DS use in individuals with obesity. We also aim to understand if demographic variables and obesity status can assist with classifying an individual’s likelihood of using any RXD or DS, not just specific RXD or DS.

Materials and Methods

The National Center for Health Statistics of CDC has been conducting the National Health and Nutrition Examination Survey (NHANES) as a continuous cross-sectional health survey. It samples the non-institutionalized population of the United States with a stratified multistage probability model and releases results of a set of health surveys, medical examinations, a
physical, and laboratory test every two years. Its rigorous quality control ensures high-quality data collection and national representativeness. The NHANES data have been used in many public health and epidemiology studies.\textsuperscript{26–32}

Demographic, physical examination, prescription drug (RXD) and dietary supplement (DS) use\textsuperscript{8}, and health insurance information were extracted from NHANES for survey years 2003 – 2014 (6 survey cycles). This is the same cycle used in a previous study\textsuperscript{33} regarding DS use which allowed for comparison. Additionally, starting in 2013-2014 ICD-10 codes were listed with the RXD giving improved information regarding RXD use. The 12-year sample weight (the number of people in the US population that a sample in the combined sample can represent) was calculated according to the analytical guideline of NHANES.\textsuperscript{25} The obese group is defined as: (1) \(\text{BMI} \geq 30 \text{ kg/m}^2\),\textsuperscript{34} (2) \(\text{age} \geq 18\).\textsuperscript{35} From the original NHANES data, 1937 respondents with no BMI and 32 respondents with no RXD use information were removed from the dataset. Table S1 lists the NHANES file, the NHANES variable names, the associated questions, and how they are referred to in this paper.

Data Analysis

Basic Characteristics: A profile for each group was created that included sex, age, race, annual household income, and health insurance status.

Statistical Analysis: We conducted multivariate logistic regression analyses to access: 1) the associations between using RXD/DS and variables of interests (i.e., demographic characteristics, poverty income ratio, insurance status, and obesity status) and 2) the associations between taking specific types of RXD and obesity status. Weighted multivariate logistic regression analyses were used to obtain odds ratios (OR) and 95% CIs with 12-year sample weight. All interested variables were introduced in the model first then backward elimination with a threshold of \(p =\)
0.05 was applied to eliminate variables. We kept only the variables that were significant in the initial model in the final model. The significance level was set as 0.05. All statistical analyses were performed by SAS software (SAS Institute Inc), version 9.4.

We performed two separate logistic regression analyses. 1) Usage based on the specific number of RXD/DS that the individual used was the dependent variable. Demographic characteristics were included as independent variables to examine whether taking a specific number of RXD/DS was significantly associated with demographic characteristics within non-obese or obese groups separately. This regression evaluated covariates down to the two groups. 2) Obesity status and demographic characteristics were included as independent variables to test their associations with taking a specific number of RXD/DS within the whole population. In the second analysis, the dichotomous dependent variable was whether participants were prescribed the specific types of RXD; obesity status was set as independent variable with reference as non-obese group.

Classification Modeling: We used Weka\textsuperscript{36} to evaluate different machine learning models for classifying whether a respondent used one or more RXD or DS, respectively. In the first round of the modelling, the variables included age group, sex, BMI category, race, annual household income, and insurance status. As we are also interested in seeing whether DS use would help classify RDX use and vice versa, in the second round of modeling, DS use or RDX use was added as a variable for classifying the use of the other type. A third round of modelling was done using the poverty income ratio (PIR) in place of the annual household income. Lastly, we further evaluated if machine learning models were able to classify how many RXD were used. For this round, we created four groups of RDX count (i.e., 0, 1-2, 3-5, >5). We used feature selection based on correlation (“CorrelationAttributeEval” in Weka) to rank the importance of the
variables. We evaluated four major machine learning algorithms including Naïve Bayes, Logistic Regression, SMO (Weka’s implementation of Support Vector Machine), and Random Forest. Deep learning techniques were not employed because of the small number of variables in this dataset. The data was preprocessed to make all numerical data nominal. 10-fold cross validation was employed. In each fold, 90% of the data was used for training and 10% of the data was used for testing. The models were compared using overall accuracy, precision, recall, F1-score, and AUROC.

Results

Basic Characteristics

Table 1 shows the basic characteristics of the two groups in regard to their RXD use. There are a few differences based on obesity status and demographics. In the non-obese group 52.64% of people report taking 1 or more RXD. The obese group has a higher reported use at 63.68%. In both groups, females report higher use than males. In addition, RXD use increases with age in both groups. Race also plays a role in reported RXD use in both groups; Non-Hispanic Whites have the highest percentages of use while Mexican Americans show the lowest percentages. Lastly, in both groups, those with health insurance reported higher RXD use than those that reported not having insurance.

Table 1. Basic characteristics of the study population

<table>
<thead>
<tr>
<th>Variable</th>
<th>Non-obese 0 RXD</th>
<th>Non-obese 1 or more RXD</th>
<th>Obese 0 RXD</th>
<th>Obese 1 or more RXD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Weight Count</td>
<td>%</td>
<td>Weight Count</td>
<td>%</td>
</tr>
<tr>
<td>Total</td>
<td>66508809</td>
<td>47.36%</td>
<td>73921759</td>
<td>52.64%</td>
</tr>
<tr>
<td>(Total sample count)</td>
<td>10903</td>
<td>49.71%</td>
<td>11090</td>
<td>50.56%</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>38107616</td>
<td>31295757</td>
<td>45.09%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>54.91%</td>
<td>45.09%</td>
<td>54.91%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14353846</td>
<td>19590434</td>
<td>42.29%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>57.71%</td>
<td>57.71%</td>
<td></td>
</tr>
<tr>
<td>Male sample count</td>
<td>6172</td>
<td>54.76%</td>
<td>5099</td>
<td>45.24%</td>
</tr>
<tr>
<td>-------------------</td>
<td>------</td>
<td>--------</td>
<td>------</td>
<td>--------</td>
</tr>
<tr>
<td>Female</td>
<td>28401193</td>
<td>39.99%</td>
<td>42626002</td>
<td>60.01%</td>
</tr>
<tr>
<td>(Female sample count)</td>
<td>4731</td>
<td>44.12%</td>
<td>5991</td>
<td>55.88%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>14836894</th>
<th>70.18%</th>
<th>6304274</th>
<th>29.82%</th>
<th>4660202</th>
<th>71.00%</th>
<th>1903380</th>
<th>29.00%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(18-24 sample count)</td>
<td>3034</td>
<td>76.69%</td>
<td>922</td>
<td>23.31%</td>
<td>948</td>
<td>74.59%</td>
<td>324</td>
<td>25.49%</td>
</tr>
<tr>
<td>25-34</td>
<td>16942291</td>
<td>65.69%</td>
<td>8849137</td>
<td>34.31%</td>
<td>7470412</td>
<td>63.24%</td>
<td>4343207</td>
<td>36.76%</td>
</tr>
<tr>
<td>(25-34 sample count)</td>
<td>2569</td>
<td>71.18%</td>
<td>1040</td>
<td>28.82%</td>
<td>1220</td>
<td>66.96%</td>
<td>602</td>
<td>33.04%</td>
</tr>
<tr>
<td>35-44</td>
<td>14731496</td>
<td>58.57%</td>
<td>10418677</td>
<td>41.43%</td>
<td>4735950</td>
<td>50.62%</td>
<td>11047591</td>
<td>49.38%</td>
</tr>
<tr>
<td>(35-44 sample count)</td>
<td>2136</td>
<td>63.89%</td>
<td>1207</td>
<td>36.11%</td>
<td>1034</td>
<td>49.38%</td>
<td>1060</td>
<td>50.62%</td>
</tr>
<tr>
<td>45-54</td>
<td>11677424</td>
<td>44.68%</td>
<td>14456138</td>
<td>55.32%</td>
<td>4735950</td>
<td>50.62%</td>
<td>10485145</td>
<td>49.38%</td>
</tr>
<tr>
<td>(45-54 sample count)</td>
<td>1546</td>
<td>48.82%</td>
<td>1621</td>
<td>51.18%</td>
<td>699</td>
<td>33.07%</td>
<td>1415</td>
<td>66.93%</td>
</tr>
<tr>
<td>55-64</td>
<td>5554943</td>
<td>29.58%</td>
<td>13224166</td>
<td>70.42%</td>
<td>2308369</td>
<td>18.04%</td>
<td>10485145</td>
<td>81.96%</td>
</tr>
<tr>
<td>(55-64 sample count)</td>
<td>942</td>
<td>33.04%</td>
<td>1909</td>
<td>66.96%</td>
<td>417</td>
<td>19.75%</td>
<td>1694</td>
<td>80.25%</td>
</tr>
<tr>
<td>65-74</td>
<td>1818148</td>
<td>14.32%</td>
<td>10874221</td>
<td>85.68%</td>
<td>570357</td>
<td>7.32%</td>
<td>7222091</td>
<td>92.68%</td>
</tr>
<tr>
<td>(65-74 sample count)</td>
<td>419</td>
<td>16.85%</td>
<td>2068</td>
<td>83.15%</td>
<td>140</td>
<td>8.82%</td>
<td>1448</td>
<td>91.18%</td>
</tr>
<tr>
<td>75 and over</td>
<td>947610</td>
<td>8.82%</td>
<td>9795144</td>
<td>91.18%</td>
<td>200980</td>
<td>5.03%</td>
<td>3795126</td>
<td>94.97%</td>
</tr>
<tr>
<td>(75 and over sample count)</td>
<td>257</td>
<td>9.96%</td>
<td>2323</td>
<td>90.04%</td>
<td>53</td>
<td>5.57%</td>
<td>899</td>
<td>94.43%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Race</th>
<th>7904327</th>
<th>72.27%</th>
<th>3033546</th>
<th>27.73%</th>
<th>4381332</th>
<th>61.83%</th>
<th>2704766</th>
<th>38.17%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mexican American</td>
<td>2359</td>
<td>65.93%</td>
<td>1219</td>
<td>34.07%</td>
<td>1197</td>
<td>54.19%</td>
<td>1012</td>
<td>45.81%</td>
</tr>
<tr>
<td>(Mexican American sample count)</td>
<td>4330888</td>
<td>63.44%</td>
<td>2496048</td>
<td>36.56%</td>
<td>2023727</td>
<td>55.70%</td>
<td>1609763</td>
<td>44.30%</td>
</tr>
<tr>
<td>Other Hispanic</td>
<td>755</td>
<td>42.85%</td>
<td>1007</td>
<td>57.15%</td>
<td>440</td>
<td>46.07%</td>
<td>515</td>
<td>53.93%</td>
</tr>
<tr>
<td>(Other Hispanic sample count)</td>
<td>39765947</td>
<td>40.57%</td>
<td>58255518</td>
<td>59.43%</td>
<td>14217576</td>
<td>29.17%</td>
<td>34528101</td>
<td>70.83%</td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>3888</td>
<td>38.14%</td>
<td>6307</td>
<td>61.86%</td>
<td>1356</td>
<td>27.06%</td>
<td>3655</td>
<td>72.94%</td>
</tr>
<tr>
<td>(Non-Hispanic White sample count)</td>
<td>7805243</td>
<td>58.33%</td>
<td>5575636</td>
<td>41.67%</td>
<td>4851478</td>
<td>42.92%</td>
<td>6452578</td>
<td>57.08%</td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>2288</td>
<td>55.11%</td>
<td>1864</td>
<td>44.89%</td>
<td>1296</td>
<td>39.58%</td>
<td>1978</td>
<td>60.42%</td>
</tr>
<tr>
<td>(Non-Hispanic Black sample count)</td>
<td>6702401</td>
<td>59.51%</td>
<td>4561010</td>
<td>40.49%</td>
<td>1285941</td>
<td>44.21%</td>
<td>1622523</td>
<td>55.79%</td>
</tr>
<tr>
<td>Other Race or Multi-Racial</td>
<td>1361</td>
<td>59.02%</td>
<td>945</td>
<td>40.98%</td>
<td>222</td>
<td>44.05%</td>
<td>282</td>
<td>55.95%</td>
</tr>
</tbody>
</table>

**Household Income**

<p>| $0 to $34,999 | 20270690 | 48.07% | 21898357 | 51.93% | 8968445 | 36.82% | 15386882 | 63.18% |</p>
<table>
<thead>
<tr>
<th>Income Level</th>
<th>Count</th>
<th>0-1</th>
<th>1.01-2</th>
<th>2.01-3</th>
<th>3.01-4</th>
<th>4.01-5</th>
<th>5.01-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0 to $34,999</td>
<td>4339</td>
<td>48.86%</td>
<td>51.14%</td>
<td>37.49%</td>
<td>62.51%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$35,000 to $74,999</td>
<td>19710293</td>
<td>47.96%</td>
<td>52.04%</td>
<td>37.40%</td>
<td>62.60%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$75,000 and over</td>
<td>21041758</td>
<td>44.89%</td>
<td>55.11%</td>
<td>33.33%</td>
<td>66.67%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poverty Income Ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-1</td>
<td>11886187</td>
<td>57.03%</td>
<td>42.97%</td>
<td>43.82%</td>
<td>66.91191</td>
<td>56.18%</td>
<td></td>
</tr>
<tr>
<td>(0-1 sample count)</td>
<td>2758</td>
<td>57.26%</td>
<td>42.74%</td>
<td>43.84%</td>
<td>1567</td>
<td>56.16%</td>
<td></td>
</tr>
<tr>
<td>1.01-2</td>
<td>13989268</td>
<td>50.62%</td>
<td>49.38%</td>
<td>39.10%</td>
<td>9996556</td>
<td>60.90%</td>
<td></td>
</tr>
<tr>
<td>(1.01-2 sample count)</td>
<td>2661</td>
<td>50.34%</td>
<td>49.66%</td>
<td>39.21%</td>
<td>1879</td>
<td>60.79%</td>
<td></td>
</tr>
<tr>
<td>2.01-3</td>
<td>9359020</td>
<td>46.75%</td>
<td>53.25%</td>
<td>34.02%</td>
<td>7567837</td>
<td>65.98%</td>
<td></td>
</tr>
<tr>
<td>(2.01-3 sample count)</td>
<td>1356</td>
<td>46.84%</td>
<td>53.16%</td>
<td>35.01%</td>
<td>1086</td>
<td>64.99%</td>
<td></td>
</tr>
<tr>
<td>3.01-4</td>
<td>8145263</td>
<td>47.14%</td>
<td>52.86%</td>
<td>35.14%</td>
<td>6173617</td>
<td>64.86%</td>
<td></td>
</tr>
<tr>
<td>(3.01-4 sample count)</td>
<td>1687</td>
<td>46.16%</td>
<td>53.84%</td>
<td>33.52%</td>
<td>1329</td>
<td>66.48%</td>
<td></td>
</tr>
<tr>
<td>4.01-5</td>
<td>16824617</td>
<td>40.22%</td>
<td>59.78%</td>
<td>30.32%</td>
<td>12809216</td>
<td>69.68%</td>
<td></td>
</tr>
<tr>
<td>(4.01-5 sample count)</td>
<td>1256</td>
<td>41.13%</td>
<td>58.87%</td>
<td>30.59%</td>
<td>878</td>
<td>69.41%</td>
<td></td>
</tr>
</tbody>
</table>

**Health Insurance**

<table>
<thead>
<tr>
<th>Insurance Status</th>
<th>Count</th>
<th>0-1</th>
<th>1.01-2</th>
<th>2.01-3</th>
<th>3.01-4</th>
<th>4.01-5</th>
<th>5.01-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>45836430</td>
<td>40.83%</td>
<td>59.17%</td>
<td>17826007</td>
<td>29.68%</td>
<td>42227916</td>
<td>70.32%</td>
</tr>
<tr>
<td>(Yes sample count)</td>
<td>6868</td>
<td>41.18%</td>
<td>58.82%</td>
<td>2783</td>
<td>29.83%</td>
<td>6546</td>
<td>70.17%</td>
</tr>
<tr>
<td>No</td>
<td>20600914</td>
<td>73.47%</td>
<td>26.53%</td>
<td>8918505</td>
<td>65.61%</td>
<td>4675065</td>
<td>34.39%</td>
</tr>
<tr>
<td>(No sample count)</td>
<td>1726</td>
<td>65.88%</td>
<td>34.12%</td>
<td>2620</td>
<td>74.56%</td>
<td>894</td>
<td>25.44%</td>
</tr>
</tbody>
</table>

*Overall insurance coverage was 80.5% covered and 19.5% not covered.

**Specific RXD Types**

For the identified RXD types, we evaluated the association between the specific types of RXD use and obesity status. Table 2 shows the odds ratio between obese and non-obese group taking prescription drugs. While looking at the types of RXD, there were differences in use between the obese and non-obese groups. Cardiovascular agents are the most used RXD type in both groups.
This is not surprising given the prevalence of cardiovascular diseases in the United States. Compared with those in the non-obese group, individuals with obesity are more likely to take cardiovascular agents (OR=1.265, 95% CI 1.222-1.311) and metabolic agents (OR=1.398, 95% CI 1.343-1.456). Figure 1 (a) shows a comparison of the cardiovascular agents and the metabolic agents based on percent of the RXD used by each age group. In both types of RXD, the obese population uses more of the RXD in the younger age groups, starting clearly at 25-34. The decrease in usage by obese individuals in the 75+ group may, in part, reflect the potential mixed effects of obesity in old age. Studies indicate that people with higher than normal BMI have lower mortality in the 75+ age group, though this is dependent on many other variables.\textsuperscript{37–39} Figure 1(b) shows the more detailed breakdown of CVD RDX use by people who are underweight, normal weight, overweight, and obese.
Table 2. Odds ratios between obese and non-obese group taking prescription drugs

<table>
<thead>
<tr>
<th>Prescription drugs</th>
<th>Obese (%)</th>
<th>Non-Obese (%)</th>
<th>Odds ratios</th>
<th>95% Wald CL</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiovascular agents</td>
<td>29.49</td>
<td>26.77</td>
<td>1.265</td>
<td>1.222</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
Correlation Analysis of RDX and DS Use with Demographic Characteristics

Examining the correlation between reported RDX use and demographic characteristics for the non-obese and obese groups, we found a few items of interest. Table S2 shows the results of this regression analysis. Male from both the non-obese and the obese group were significantly less likely than female to use RDX (OR<sub>control</sub> = 0.572, 95% CI 0.538-0.607, OR<sub>obese</sub> = 0.552, 95% CI 0.508-0.626). Compared with those individuals with obesity older than 75, adults younger than 54 were significantly less likely to use RDX. When controlling for all other variables, non-obese people covered by insurance were around 1.338 times (p<.0001 for non-obese group and p=0.0056 for obese group) as likely to use RDX than those who did not have any insurance coverage. Specifically, individuals with obesity covered by Medicare were 1.435 times more likely to use RDX (p<.0001) than those with no Medicare covered.

When looking at the correlation between reported DS use and demographic characteristics for the non-obese and obese groups, we found a few interesting findings (Table S3). Within the obese group, Mexican American, other Hispanic, and Non-Hispanic Black, and Other Races (including multi-racial) were significantly less likely to take DS compared with Non-Hispanic White. A significant correlation was found between PIR (Poverty income ratio: a ratio of family income to poverty threshold) and DS use for the control group. Non-obese people...
with a PIR higher than 4 are more likely to take DS. Individuals with obesity covered by private
insurance, Medicare, and other government insurance were significantly more likely to take DS,
while individuals without obesity covered by insurance were 1.241 times as likely to take DS
than those who were not covered by any insurance while holding other variables constant.

We were also interested in how obesity status and demographic characteristics associated
with the use RXD or that of DS within the whole population group (Table 3). Female and older
people were more likely to take RXD or DS. PIR is an interesting factor, as people with higher
PIR were significantly more likely to take DS. For RXD use, only those with PIR >= 5 were
found to be statistically significant. This was also the case with the PIR (Poverty income ratio: a
ratio of family income to poverty threshold) showing that those with a PIR higher than 1 are
more likely to take RXD or DS. The higher the PIR is, the higher odds people use DS.
Individuals with obesity were more likely to take RXD (OR= 1.567, 95% CI 1.481-1.658) while
less likely to take DS (OR= 0.769, 95% CI 0.731-0.81) compared with individuals without
obesity.

Table 3. Reported prescription drug use/dietary supplements use by demographic characteristics
and obesity status correlates.

<table>
<thead>
<tr>
<th>Variable</th>
<th>RXD use</th>
<th>DS use</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds ratio</td>
<td>95% Wald CL</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.549</td>
<td>0.521</td>
</tr>
<tr>
<td>Female (reference)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Age Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>0.105</td>
<td>0.084</td>
</tr>
<tr>
<td>25-34</td>
<td>0.132</td>
<td>0.107</td>
</tr>
<tr>
<td>35-44</td>
<td>0.198</td>
<td>0.16</td>
</tr>
<tr>
<td>45-54</td>
<td>0.335</td>
<td>0.27</td>
</tr>
<tr>
<td>55-64</td>
<td>0.618</td>
<td>0.497</td>
</tr>
<tr>
<td>65-74</td>
<td>0.693</td>
<td>0.565</td>
</tr>
</tbody>
</table>
Figure 2 shows the correlation between age, BMI, and the number of RDX and DS used by both the obese group and non-obese group. Figure 2(a) illustrates the correlation between the average number of RXD used by a respondent and the age groups. Generally, the average
number of RXD used by a respondent increased with an increase in age for both non-obese and obese groups. Regardless of age, individuals with obesity generally take more RXD than individuals without obesity. The distribution of the average number of RXD used shows a positive skewness distribution: the average number is greater than the median within each age group. Figure 2(b) illustrates the correlation between the average number of DS used and the age groups. Similarly, the average number of DS used generally increased with an increase in age for both groups, with people aging from 65 to 74 taking the highest number of DS (mean of non-obese vs. obese: 2.28 vs. 1.86). However, the difference in average number of DS used between the obese and non-obese groups was not as clear as that between the number of RXD used and age group. Figure 1(c) illustrates the correlation between the number of RXD/DS used and BMI. BMI mainly clustered around 18 to 37 kg/m². People typically take higher numbers of RXD than DS. The largest maximum number of DS used (24) can be found among people with BMI from 18 to 22 kg/m². The highest average number of RXD used can be found among people with BMI from 63 to 67 kg/m². As both of the BMI categories from 83 to 87 kg/m² and from 128 to 132 kg/m² have only one person included in the sample, no bar shows in the figure, the points indicate the true value among those groups.
Figure 2. (a) The correlation between the number of RXD used and age; (b) The correlation between the number of DS used and age; (c) The number of RXD/DS used compared to BMI. The shade of color represents the aggregated weights within each age group. The points of line indicate the average number used within each age group. The shade of the color represents the
aggregated weights within each age group. The points on the line indicate the average number used within each age group.

**Classification of RDX Use and DS Use Using Machine Learning**

Table S4, S5, S6, S7, and S8 in the Supplementary Material show the detailed results of classifying DS and RXD use. Classification of DS use (binary variable) was not as accurate as classification of RXD use (binary variable). Results from the models run were similar for DS use with the highest overall accuracy being 64.63% and the AUROC at 0.7 (Table S4). The results for RXD use were better, with the highest overall accuracy being 74.27% and the highest AUROC at 0.816 (Table S5). To see if results would change with the addition of the variable DS use (for RXD use classification) and RXD use (for DS classification) the same machine learning algorithms were evaluated again. The overall accuracy and AUROC increased slightly in both DS models (Table S4) and RXD models (Table S5). This shows that adding the extra variable did not make a significant contribution to the classifications. Using the PIR in place of the annual household income also did not significantly change the results for classification of both DS use (Table S6) and RDX use (Table S7). For DS classification, RXD use, insurance status, and sex were the top three important features. For RXD classification, insurance status, DS use, and age were the top three important features. In both cases, BMI category and annual household income were in the least three important features. We further created four groups for RDX count (i.e., 0, 1-2, 3-5, >5) and classified samples into one of these groups, using the same models and variables. In this experiment (Table S8), the best model is logistic regression with an AUROC of 0.76. Note that multiclass classification is inherently more challenging than binary classification.
Discussion

Individuals with obesity experience increased risk for developing chronic diseases across the lifespan, which are often managed using RXD. Overall, RXD use in the United States has increased, with the greatest increases seen in the treatment of conditions found to be associated with obesity, specifically antihypertensives, antihyperlipidemic, antidiabetics, and antidepressants. Further, many individuals may use DS in addition to or instead of RXD. Given the increases in obesity, RXD use, and DS use, it is important to analyze trends in order to better characterize individuals RXD and DS use in relation to obesity.

In this study, we used NHANES data from 2003-2014 to examine RXD and DS use in relation to obesity status. We showed that demographics and obesity status do play a role in usage. In regard to demographic variables, we found that the obese group has a higher reported use of RXD at 63.68%. In both groups, females report higher use than males. In addition, RXD use increases with age in both groups. The difference in usage based on sex is explainable because females tend to have more consistent visits to medical practitioners and typically use more RXD in general, compared to males. The increase in RXD use shown with age can be explained by increased prevalence in clinical comorbidities as the population gets older and is consistent with prior population studies. The reason why the RXD usage decreases by obese individuals in the 75+ group may be that individuals whose obesity was associated with CVD earlier in life may have higher mortality rates. Another possibility is that higher BMI in older age may be protective, as previous research has suggested. Regardless, our findings further highlight that BMI is not as useful of a health parameter in older adults as it is in young and middle-aged adults, which is consistent with previous research. The differences in race can be explained by the healthcare and insurance gap seen in minority races. The increased use by those that are...
insured can be explained by an increased use of health care services and the associated increase
coverage of RXD.\textsuperscript{44}

Regarding specific RXD types, cardiovascular agents and metabolic agents were used more by the obese group, while hormone/hormone modifiers and psychotherapeutic RXD usage was higher in the non-obese group. While increased use of cardiovascular and metabolic RXD was expected given the cardiometabolic comorbidities with obesity, higher usage of hormonal and psychotherapeutic RXD in non-obese individuals was surprising.\textsuperscript{44} Prior research has found higher rates of reproductive issues and increases in depression and other mental health disorders associated with obesity;\textsuperscript{45} thus, we would expect higher use of related RXD. One explanation may be the association between high TSH and high BMI and low free-T4 and high BMI. These levels may not be outside of the “normal” range for these values but still cause an increase in BMI which could mean that RXD use would not necessarily be indicated.\textsuperscript{46} In regard to psychotherapeutic RXD, it is possible that individuals with obesity may have unrecognized or undiagnosed mental health issues that are seen as medically related to obesity rather than mental health. For example, unhealthy coping mechanisms, such as increasing food intake or binging may be related to undiagnosed depressive symptomatology.\textsuperscript{47}

In looking at tracking the use of specific RXD prescribed for weight loss, there was a small proportion of the population that utilized these drugs. The problem with further study of these RXD is that many of them were approved outside of the 12-year survey data used in this project. Additionally, many of the drugs used for weight loss are also used for other purposes such as management of diabetes or as a general CNS stimulant. Based on the information available it was unclear why respondents used a specific RXD. This made it difficult to understand and analyze the use of these drugs.
When looking at both DS use and RXD use, the population with obesity was more likely to use RXD, but less likely to use DS compared with the non-obesity population. This finding is consistent with a previous report that only 33.9% of adults use DS for weight loss. Similar results were also obtained regardless of sex, age, and financial status. However, we found additional novel predictors, such as insurance status. While higher RXD use in obese individuals was expected, lower use of DS was surprising. Given the low percentage of individuals who use DS for weight loss, it is possible that a majority of DS use is related to seeking other purported benefits. For example, there are a plethora of DS marketed toward enhancing brain function or enhancing physical function, which may be more likely to be consumed by either older or non-obese individuals who are already health-conscious.

The experiments with machine learning models showed that predictions of RXD and DS use could be improved. Future studies should examine alternative variables that may better predict RXD and DS use. For example, diagnostic information, out-of-pocket costs, and RXD coverage information may provide valuable additional information for understanding these relationships. Validated models of RXD and DS use would provide valuable information that may inform patient education. Further, physicians and other health care providers may be able to use this information to better understand patient trends and to make informed prescribing decisions. In particular, health care providers may benefit from being able to characterize individuals most likely to use DS, as these may go unreported otherwise.

Overall, our findings indicate that obesity is associated with higher RXD and lower DS use, but income and age are important demographic factors to consider in this relationship. Since it appears that the more used RXD types are associated with comorbid conditions related to obesity instead of treating obesity directly, there may be opportunities for better health
education. This may include expanding education on the benefits of lifestyle changes to minimize both obesity and the impact of common comorbid conditions. Further, the lower observed use of DS in the obese population may offer an opportunity for patient education on DS like omega-3 fatty acids and Vitamin B/B-Complex, which are shown to benefit cardiovascular health, omega-3 fatty acids which are shown to benefit weight loss and to maintain blood sugar, and multivitamin/multiminerals which are shown to impact general overall health. Understanding use patterns of DS in obese and non-obese individuals may also provide opportunities for educating patients about potentially harmful or ineffective DS, such as those targeting weight loss specifically.

Based on Figures 1 and 2(a), there seems to be a point in the 20-30 age range that would be ideal for addressing obesity proactively and aggressively before it escalates into having comorbidities that require RXD. Additionally, there is an opportunity for better education in those under 20 as studies show that obesity in childhood and adolescence can lead to obesity as an adult. Education has shifted from education on treating obesity to education on prevention of obesity, given that losing small amounts of weight or maintaining a healthy weight are more effective than treating obesity once it has developed. Future studies should examine how access to healthcare may further impact these relationships, especially in regards to accessing quality obesity prevention education.

**Limitations**

In most of the survey cycles used, there was no reason or diagnosis code associated with the report of RXD use. This means that we do not know why a respondent was taking a certain RXD. In addition, only primary category was used to type the various RXD. This means that if an RXD has multiple uses or off-label uses that cross-disease categories, it would not be evident.
in these results. Further, NHANES does not have information specifically geared towards
delineating what may be considered a DS whose main purpose is weight loss. Another limitation
is that NHANES is a cross-section survey. It is thus not possible to infer any casual relationships
between the variables.

**Future Work**

The most recent release of NHANES survey cycle (2017-2018) does include a diagnosis code
with the RXD information. A future project that contains richer data in regard to RXD and DS
use would provide more insight. Even though national health surveys like NHANES provide
enormous opportunities for answering many important health-related questions, they have not
been used widely for patient education and informatics research except for a few studies of our
own.\textsuperscript{53-55} In future work, we plan to build informatics tools such as a data dashboard to visualize
various types of analyses of the NHANES data to provide patients, policy makers, and health
providers a way to explore RXD and DS use in the general population and certain population
subgroups.

**Conclusions**

As obesity becomes a larger issue and the weight crisis in the United States becomes
increasingly detrimental, more needs to be done to understand the overall health status associated
with this population and how to educate the public about obesity, its comorbidities, and
preventive measures that may help. Knowing how RXD and DS use are different from those
without obesity is only the start. Further steps can be taken to understand why there are
differences and how the underlying diseases and conditions can be pre-empted in this group.
Further knowledge on the association between obesity, DS, and RXD can inform patient
education with the help of informatics tools such as data dashboards. Developing models that
help us understand the causes of and lifestyle changes needed to change obesity status should improve overall health in the United States.

Competing Interests

None

References


Figure Caption

Figure 1. (a) Comparison of RXD used by age group. CVD is cardiovascular drug. MA is metabolic agent. (b) Comparison of CVD RXD used by different age groups and weight categories.

Figure 2. (a) The correlation between the number of RXD used and age; (b) The correlation between the number of DS used and age; (c) the number of RXD/DS used compared to BMI. The shade of color represents the aggregated weights within each age group. The points of line indicate the average number used within each age group. The shade of the color represents the aggregated weights within each age group. The points on the line indicate the average number used within each age group.

Table Caption

Table 1. Basic characteristics of the study population

Table 2. Odds ratios between obese and non-obese group taking prescription drugs

Table 3. Reported prescription drug use/dietary supplements use by demographic characteristics and obesity status correlates.

Supporting Information

Table S1. NHANES Variables

Table S2. Reported prescription drug use by demographic characteristics among obese and control group.

Table S3. Reported prescription dietary supplements use by demographic characteristics among obese and control group.

Table S4. Performance of machine learning models for classifying DS use

Table S5. Performance of machine learning models for classifying RXD use

Table S6. Performance of machine learning models for classifying DS use using PIR

Table S7. Performance of machine learning models for classifying RXD use using PIR

Table S8. Performance of machine learning models for classifying RXD use into categories