DATA QUEST – *Unlocking the Power of Data*

Case Study Challenge 2023

Statistics Track – Final Round

From fork to fitness!

Exploring the Relationship Between Eating Patterns and Obesity

Ayshik Neogi
Darshan Mali
Saurav Jadhav

Savitribai Phule Pune University

Idea Generation

1. We generated all the numerical variables from the normal distribution according to the parameters specified and the proportions given for each treatment.
2. For the questionnaire, we used the truncated normal distribution.
3. We made the values for BMI and body weight missing according to the missing data proportion.
4. The countries, sex, and the age variable values are uniformly sampled.
5. Imputation of missing values are carried out using the KNN method.

Data Summary

1. The majority of the teens under study are from countries with a low economy.

2. The empirical distribution of BMI, and weight of the teens are multimodal which indicates that it is more likely we have groups and subgroups within the data based on different characteristics.

3. The added sugar consumption decreases with the visits for all the treatments including placebo.

4. From all the numerical variables under consideration, ECS is the only one which shares a relatively stronger relationship with multiple variables: RCS, Added_sugar, UCS and BMI.

Exploratory Analysis

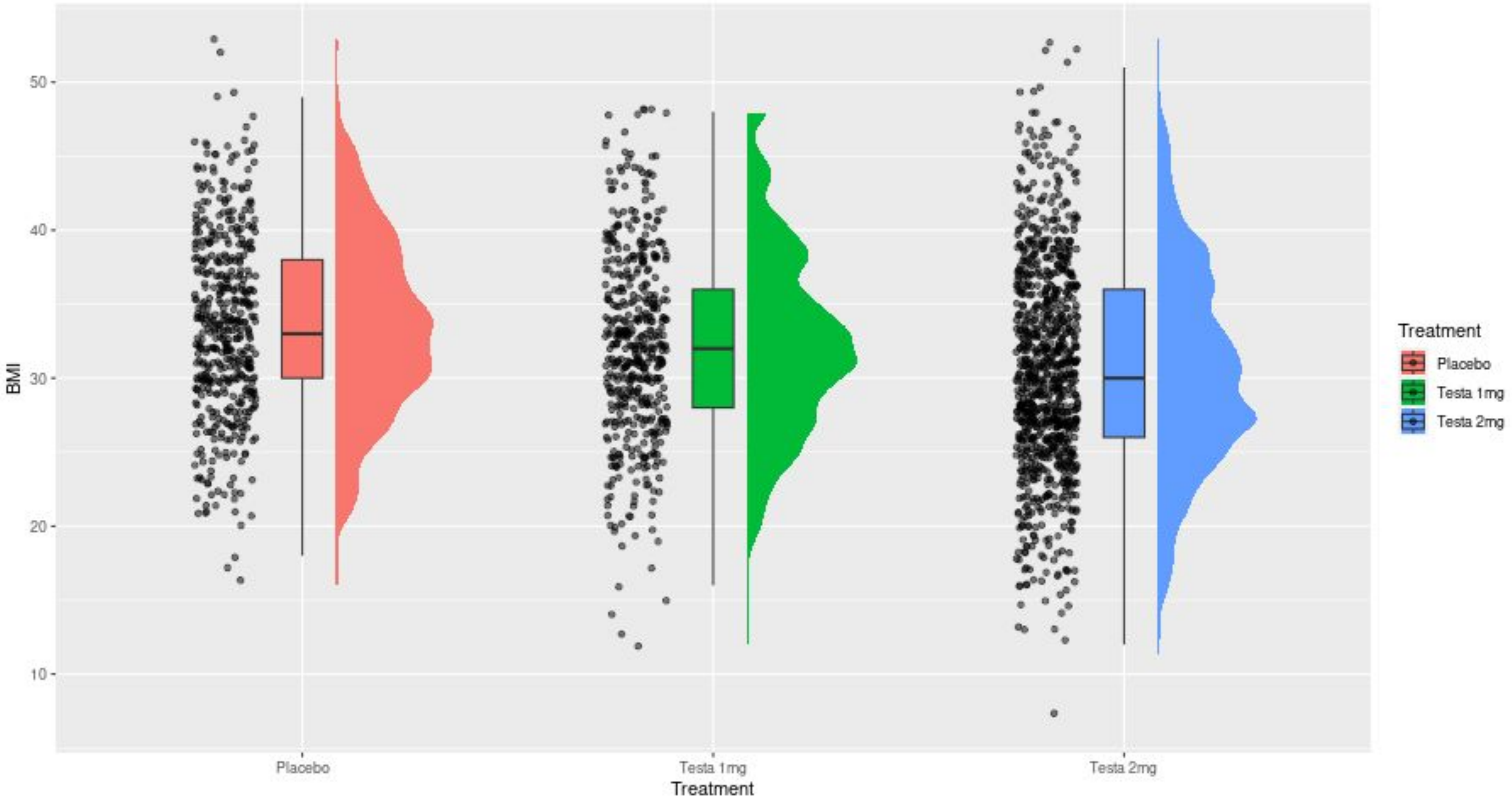

Fig 1: Density plot of BMI for all the treatment type. From the plot we can observe multiple modes for different treatment, and we can also observe the different spread.

Questionnaire based analysis

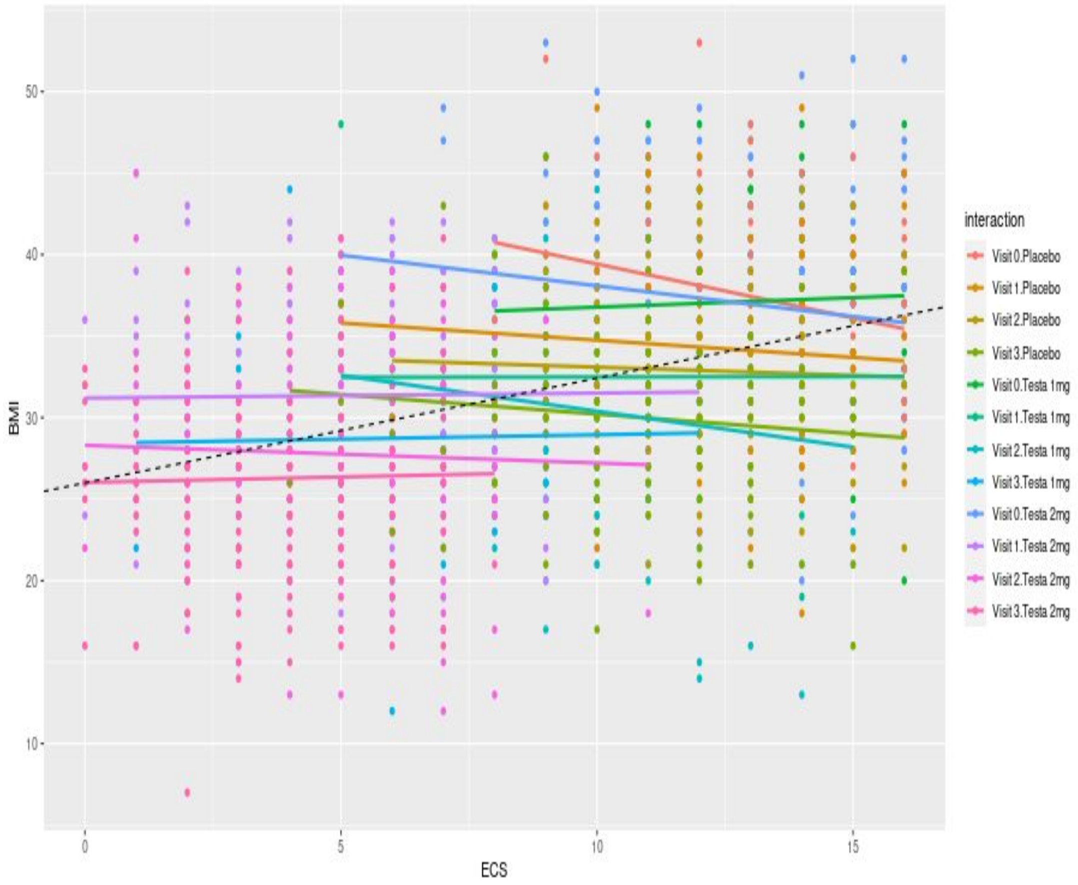

Fig 2: Simpson's Paradox on BMI vs ECS with Treatment & Visits as confounders

1. We used MI (mutual Information) to observe which relationships are significant.
2. We found ECS and BMI related.
3. After introducing the confounders, that is, treatment:visits, we observe no relationship between BMI and ECS.
4. We found no significant relationship between TestQ and Impulsivity Questionnaire.

Primary endpoint analysis-1

Visit 0					
Contrast	estimate	SE	df	t.ratio	p.value
Placebo - Testa 1mg	0.117	0.727	1904	0.161	0.9859
Placebo - Testa 2mg	0.221	0.63	1904	0.35	0.9346
Testa 1mg - Testa 2mg	0.104	0.63	1904	0.165	0.9851
Visit 1					
Contrast	estimate	SE	df	t.ratio	p.value
Placebo - Testa 1mg	1.675	0.727	1904	2.305	0.0553
Placebo - Testa 2mg	2.815	0.63	1904	4.47	<.0001
Testa 1mg - Testa 2mg	1.14	0.63	1904	1.811	0.1664
Visit 2					
Contrast	estimate	SE	df	t.ratio	p.value
Placebo - Testa 1mg	2.617	0.727	1904	3.601	0.0009
Placebo - Testa 2mg	5.138	0.63	1904	8.16	<.0001
Testa 1mg - Testa 2mg	2.522	0.63	1904	4.005	0.0002
Visit 3					
Contrast	estimate	SE	df	t.ratio	p.value
Placebo - Testa 1mg	1.267	0.727	1904	1.743	0.1895
Placebo - Testa 2mg	3.741	0.63	1904	5.94	<.0001
Testa 1mg - Testa 2mg	2.474	0.63	1904	3.928	0.0003

Table 1

1. We use mixed effect model with subject ID as random effect, and visits and treatment as fixed effect. The interaction of treatment and visit is also considered.

2. We use emmeans() to estimate the marginal difference between multiple treatments, adjusted for the different effects in the model.

3. We can see the difference for the Testa 1mg and Testa 2mg is positive for all the visits, and it is significant for the visit 3 at 5% level of significance.

Primary endpoint analysis-2

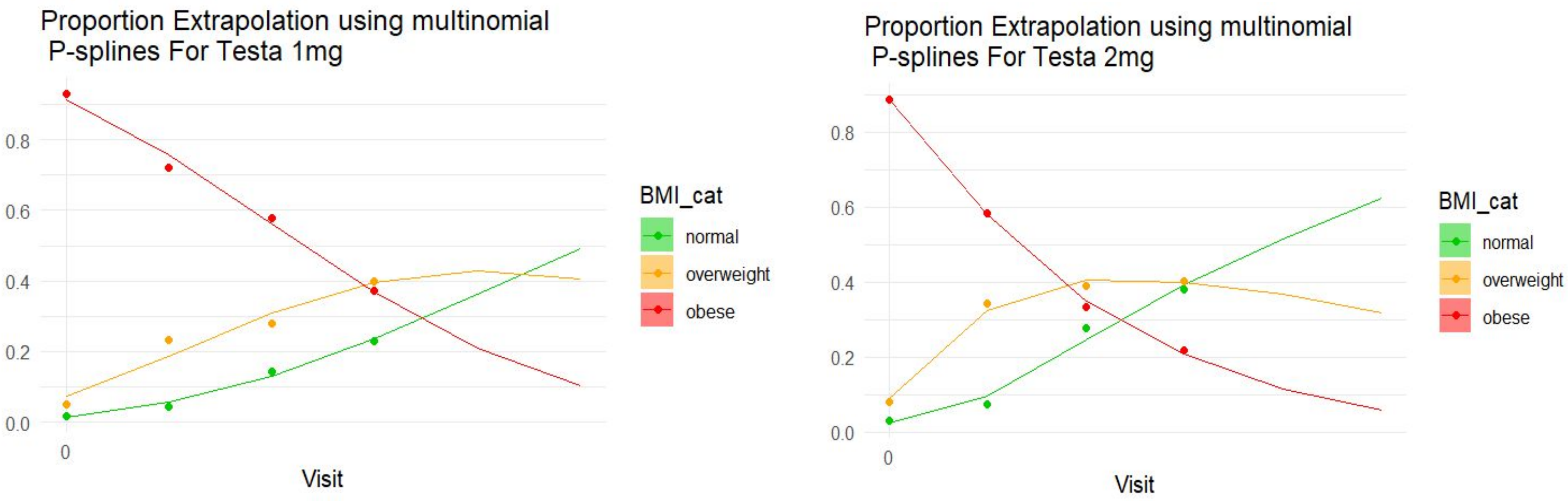

Figure 3

In order to estimate the net transition probabilities, we need to have smooth estimates of the Visit-specific prevalences for each BMI category. Each BMI category gets its own P-spline, because the third category is one minus the sum of the others. Extrapolation results (Figure 2) over visit clearly says that Testa 2mg is more effective than Testa1mg.

Primary endpoint analysis-3 [1.3]

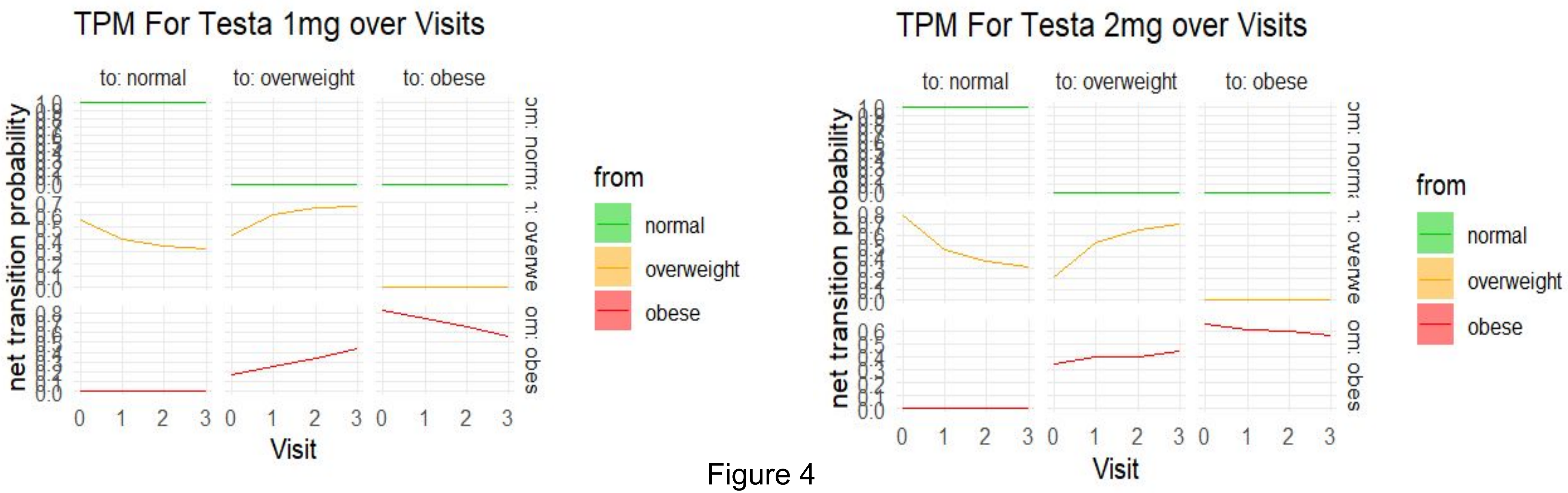

Figure 4

We can get Estimated value for Transition probabilities from multinomial P-spline fit through observed prevalence proportions of weight category.

1. If our target subject is overweight then Testa 1 mg shows better results than Testa 2mg as the probabilities from being overweight to normal is higher in Testa 1 mg.
2. However if target subject is obese then Testa 2mg is more effective as probability of transition from being obese to overweight is high in Testa 2mg than Testa 1 mg.
3. So, we can use combination of Testa 2mg(Obese to Overweight) and Testa 1mg(Overweight to Normal) for transforming obese subject to normal.

Primary endpoint analysis for [1.4]

Treatment	Visit	emmean	SE	df	lower.CL	upper.CL
Placebo	Visit 0	37.3	0.514	1904	36.3	38.3
Testa 1mg	Visit 0	37.2	0.514	1904	36.2	38.2
Testa 2mg	Visit 0	37.1	0.364	1904	36.3	37.8
Placebo	Visit 1	34.2	0.514	1904	33.2	35.2
Testa 1mg	Visit 1	32.5	0.514	1904	31.5	33.5
Testa 2mg	Visit 1	31.4	0.364	1904	30.6	32.1
Placebo	Visit 2	32.9	0.514	1904	31.9	33.9
Testa 1mg	Visit 2	30.3	0.514	1904	29.3	31.3
Testa 2mg	Visit 2	27.8	0.364	1904	27	28.5
Placebo	Visit 3	30	0.514	1904	29	31
Testa 1mg	Visit 3	28.8	0.514	1904	27.7	29.8
Testa 2mg	Visit 3	26.3	0.364	1904	25.6	27

Table 2

1. Total 363 subjects are qualified for the study of [1.4].
2. Similar to the approach adapted for [1.3], we use the mixed effect model.
3. We use emmeans() to calculate the marginal mean adjusted for the different effects in the model.
4. For this case, we also compute the 95% confidence interval for BMI.
5. We observe, Testa 2mg has the confidence interval with lowest BMI values.

Primary endpoint analysis for [1.4]-2

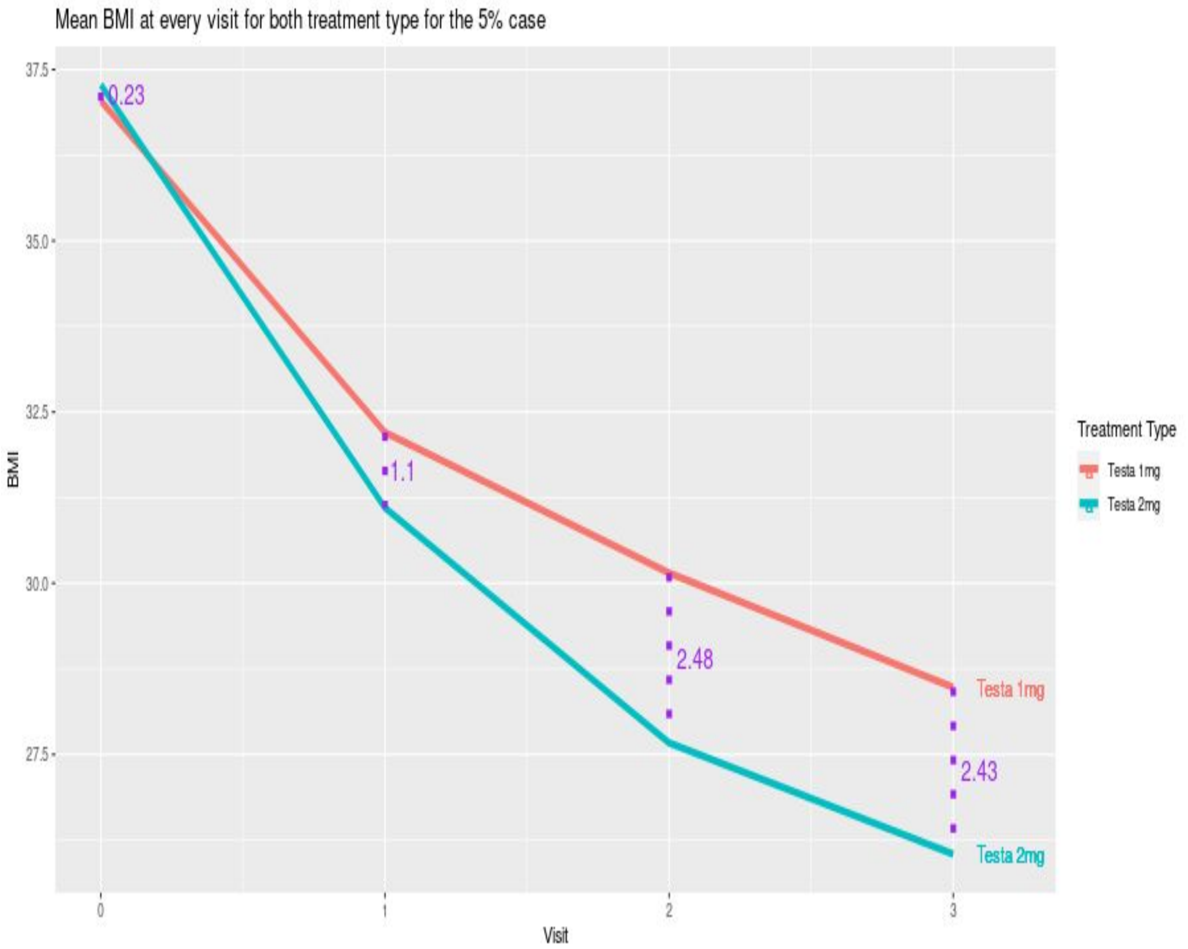

Figure 4

1. To decide which treatment is more effective, we use a unique approach. At every visit, we consider the mean BMI for all those teens who were given testa 1mg. We do a similar thing for those who were given testa 2mg.
2. Now we calculate the distance between both the treatment means of the BMI for every visit. We compare the distance of the last visit to that of the first visit.
3. We observe that, even if both the treatment shows almost similar average BMI initially, there's a larger fall in the BMI for those teens who were given testa 2mg.

Consistency & a Single Informative Score

1. We use Cronbach's alpha to measure the internal consistency of the scores. We found the cronbach's alpha to be 0.522
2. To design a single informative score, we first standardized all the components of TestQ and Impulsivity Questionnaire.
3. We multiplied the standardized RCS by negative sign so that, for all the questionnaire, the lower value will indicate favorable response.
4. Now we sum the standardized response of all these components to have a single score.
5. We found out that the single score has a significant positive relationship with BMI.

Sugar Tax Policy

1. Originally, the sugar tax policy was implemented in Hungary and was studied by Biro (2015)

2. The changes in the relative prices of unhealthy and healthy foods and beverages can lead to significant dietary improvements and weight losses, particularly among those who are most at risk of obesity. Powell et al. (2013)

3. The household consumption of sugar-sweetened beverages increases with household income in the early stage, but then tends to fall with income growth after reaching the turning point (at the household income of 400,300 yuan). Liu, Z., Li, S., & Peng, J. (2022)

4. The tax policy under study will have a greater impact on SSB prices.

5. From the variable, added_sugar_consumption_per_day, we classified the teens into groups: High_tax, middle_tax, and low_tax.

6. We used mixed effect model to see whether these categories share any relationship with BMI, that is, to see the impact of the tax policy. We have observed no significant relationship between tax group and BMI.

Recommendations/Insights/Scope

1. The index that we created can be used to understand the psychological behaviour that is related with obesity and help us address the problem from its roots.

2. All the methods used to compare the two treatments reflects that Testa 2mg is more effective and we can go further to study the side effects as well as the long term effect of this drug.

3. Basu et al [2014] suggests a 20% tax rate on soft drinks, that may lead to a reduction of 3% in obesity. A 20% tax rate is relatively a smaller tax burden compared to the sugar tax policy.

4. It is recommended that the policy is rolled out in parts, which will help better understand the unseen consequences.

5. ECS showing relatively strong relationship with multiple variables can be studied further to better understand relation of emotions with other attribute, which can help us tackle obesity.

6. Depending on the weight category, we can select the treatment to reduce obesity.

References-1

1. Basu S., Vellakkal S., Agrawal S., Stuckler D., Popkin B., Ebrahim S. Averting obesity and type 2 diabetes in India through sugar-sweetened beverage taxation: An economic-epidemiologic modeling study. PLoS Med. 2014;11 doi: 10.1371/journal.pmed.1001582.

2. Pfinder, M., Heise, T. L., Boon, M. H., Pega, F., Fenton, C., Griebler, U., Gartlehner, G., Sommer, I., Katikireddi, S. V., & Lhachimi, S. K. (2020). Taxation of unprocessed sugar or sugar-added foods for reducing their consumption and preventing obesity or other adverse health outcomes. The Cochrane Database of Systematic Reviews, 2020(4). https://doi.org/10.1002/14651858.CD012333.pub2

3. Katzmarzyk PT, Broyles ST, Champagne CM, Chaput JP, Fogelholm M, Hu G, Kuriyan R, Kurpad A, Lambert EV, Maia J, Matsudo V, Olds T, Onywera V, Sarmiento OL, Standage M, Tremblay MS, Tudor-Locke C, Zhao P. Relationship between Soft Drink Consumption and Obesity in 9-11 Years Old Children in a Multi-National Study. Nutrients. 2016 Nov 30;8(12):770. doi: 10.3390/nu8120770. PMID: 27916866; PMCID: PMC5188425. 2016

References-2

4. Liu, Z., Li, S., & Peng, J. (2022). Exploring the Relationship between Sugar and Sugar Substitutes—Analysis of Income Level and Beverage Consumption Market Pattern Based on the Perspective of Healthy China. Nutrients, 14(21). https://doi.org/10.3390/nu14214474

5. Anikó Bíró, Did the junk food tax make the Hungarians eat healthier?, Food Policy, Volume 54, 2015, Pages 107-115, ISSN 0306-9192, https://doi.org/10.1016/j.foodpol.2015.05.003.

6. Kassteele Jv, Hoogenveen RT, Engelfriet PM, Baal PH, Boshuizen HC. Estimating net transition probabilities from cross-sectional data with application to risk factors in chronic disease modeling. Stat Med. 2012 Mar 15;31(6):533-43. doi: 10.1002/sim.4423. Epub 2011 Dec 5. PMID: 22139860.https://pubmed.ncbi.nlm.nih.gov/22139860/