For this analysis, I will be using data from the DEXASLIM study, which was a wait list control study assessing the effectiveness of an Artificial Intelligence (AI) health program called DEXASLIM. The scientific questions of interest for this analysis are: Is there evidence that people who participated in this trial lost weight over the 6-week follow-up period in the absence of the DEXASLIM intervention? Is there evidence that the DEXASLIM intervention is effective? Does this depend on the sex of the participant?

For the descriptive analysis section of this report, I will create several visualizations to describe the data and determine potential future steps for the analysis. Through these descriptive figures and tables I will effectively assess the population of interest, potential outliers, missing data, effect modifiers/confounders, and the distribution of the outcome of interest. These visualizations will help inform the analytic part of this report and confirm or deny assumptions about the data. To describe the demographics of the study population and check for outliers or missing data, I will create a summary table, a double boxplot, and an overlayed/double histogram.

The summary table will include the title: “Table 1: Demographics of study participants by intervention status (n=54)”. The rows in the table will be the collected variables of *sex, preweight & week6weight*. The *sex* variable is dichotomous categorical, so for this variable I will have 2 rows - one for Male and one for Female - and report the number of participants in each category as well as the proportion (%). The *preweight* and *week6weight* variables are continuous numerical variables, so I will record the mean and standard deviation for each. Additionally, I will create a new variable, *weightdiff*, that will calculate the difference between the *preweight* and *week6weight* measurement. This will also be reported as a mean and standard deviation. Lastly, in order to assess these aforementioned variables in the context of the scientific question - Is there evidence that the DEXASLIM intervention is effective? - I will stratify the above rows by intervention status, which is the variable called *group*. This will allow us to effectively describe the study population, as well as the reported demographics based on the intervention status. It will also allow us to see if there are any missing data across the different variables.

The double boxplot will include the title: “Figure 1: Boxplots of intervention change in weight by sex (n=27)”. This boxplot will aim to assess whether there is a difference in weight change (using the aforementioned *weightdiff* variable that I created) based on the *sex* variable. There will be 2 boxplots side-by-side, showing the median, IQR and any outliers across the *weightdiff* variable, stratified by Male and Female. By comparing the distribution of data points in the boxplots, we can determine if there is effect modification or confounding happening with regards to sex in this study, addressing the final scientific question of interest.

**Null hypothesis:** There is no difference in the relationship between DEXASLIM and weight loss based on sex. (μ1 = μ2)

**Alternative hypothesis:** There is a difference in the relationship between DEXASLIM and weight loss based on sex. (μ1 ≠ μ2)

The main assumption of this analysis is that the data are normally distributed. This will be accomplished through visualizing the data and assessing their distribution. Based on the hypothesis test that I intend to perform (two sample t-test) it must also be assumed that the variances of both samples are equal. Another assumption (which cannot be statistically tested but must be noted) is that the data were collected systematically across both groups and the trial was conducted as stated.

I will use a double histogram to ensure that the data are normally distributed. I must stratify the data and do two histograms instead of one since there are two independent sample groups (intervention and waitlist control). The overlayed/double histogram will be titled “Figure 2: Distribution of the change in weight by intervention status (n=54)”. Since the sample size is small when the sample is stratified by intervention, it will be important to assess the distribution of the data to ensure that a parametric hypothesis test is appropriate. If the distribution is approximately normal for both groups, then I know that a parametric hypothesis test is possible and appropriate (and will provide greater statistical power than a non-parametric test).

Given that the data are approximately normally distributed based on the assessment of the double histogram, I will be moving forward with a parametric hypothesis test to address the second scientific question: Is there evidence that the DEXASLIM intervention is effective? My hypothesis is that the DEXASLIM intervention group will have greater weight loss compared to the waitlist control group. I will assess this with a two sample t-test, since it allows for comparison between two independent samples. After calculating the weight change within each group (*weightdiff*) I will find the mean value for each group and their standard deviation. I will use the t.test function in R to calculate the test statistic (t) along with the p-value and 95% confidence interval. If the p-value is less than α = 0.05 and the confidence interval does not cross 0, then we can reject the null hypothesis. Otherwise, we would fail to reject the null hypothesis.

**Null hypothesis:** There is no difference in weight loss between the DEXASLIM group and the control group. (μ1 = μ2)

**Alternative hypothesis:** There will be greater weight loss in the DEXASLIM group then in the control group. (μ1 < μ2)

To assess the final scientific question of interest, Is there evidence that people who participated in this trial lost weight over the 6-week follow-up period in the absence of the DEXASLIM intervention? I will use a paired t-test to see if there is a difference in weight from the beginning and the end of the collection period for the waitlist control group. I will use the t.test function in R to calculate the test statistic (t) along with the p-value and 95% confidence interval. If the p-value is less than α = 0.05 and the confidence interval does not cross 0, then we can reject the null hypothesis. Otherwise, we would fail to reject the null hypothesis.

**Null hypothesis:** There is no weight loss in the waitlist control group. (μ1 = μ2)

**Alternative hypothesis:** There is weight loss in the waitlist control group. (μ1 > μ2)

**Appendix 1 – Variables table**

| **Name** | **Description** | **Coding or Units** | **Additional comments (e.g., data type, variable type, range)** |
| --- | --- | --- | --- |
| weightdiff | week6weight - preweight | kg | The difference in weight from week 0 to week 6 |
| sex.f | Sex (string) | 0 = “female”, 1 = “male” | Taken from sex variable |
| group.f | Group (string) | 0= “control”, 1 = “intervention” | Taken from group variable |

**Appendix 2 - Mock Tables and Figures:**

Table 1: Demographics of study participants by intervention status (n=54)

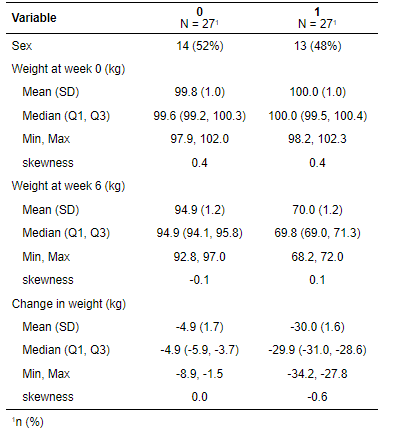


Figure 1: Boxplots of intervention change in weight by sex (n=27)

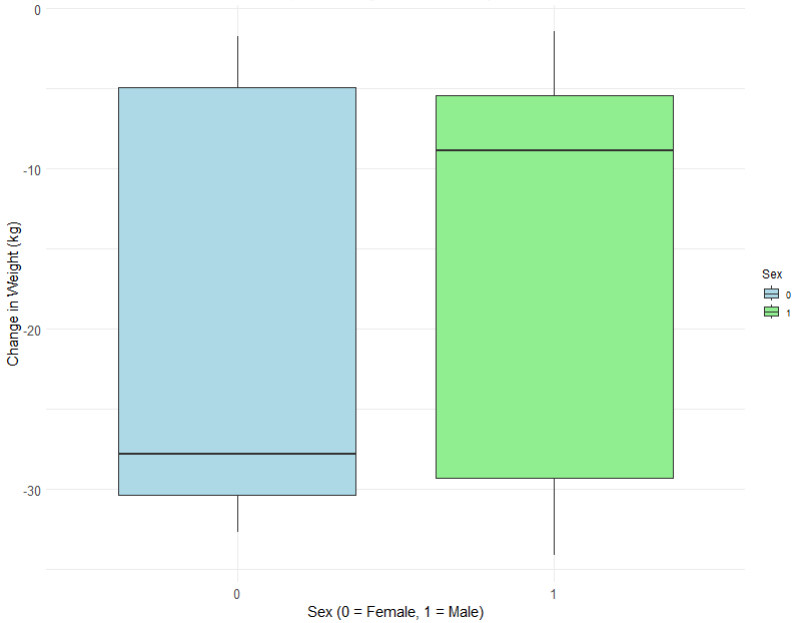


Figure 2: Distribution of the change in weight by intervention status (n=54)

