Title: Analyzing the Potential of Miraculon-B: A Groundbreaking Advancement in Solid Tumor Treatment

Introduction:

Developing effective treatments for solid tumors remains a significant challenge in the field of oncology. Researchers at one of the top global biopharma companies with presence in the UK, US, Europe and Asia have been working tirelessly to address this problem by introducing an innovative family of drugs. Among these potential treatments, Miraculon-B (a pseudonym) has shown promising results and is currently undergoing evaluation in late-phase clinical trials. This trial aims to determine whether Miraculon-B is more effective than the standard of care in shrinking solid tumors for patients who do not respond to other treatment options.

As the clinical trial reaches its conclusion, the next crucial step is to analyze the data obtained during the study. The analysis will provide insights into the effectiveness of Miraculon-B compared to the standard of care, as well as identify patient sub-groups that may benefit more from the new drug. This information is vital for the development of a comprehensive package to be presented to regulatory agencies, which will outline the likelihood of Miraculon-B's success and propose a strategy for its prescription to healthcare providers.

Problem Statement:

The objective of this data analysis project is to evaluate the effectiveness of a new cancer treatment drug called Miraculon-B. The drug is targeted towards patients with solid tumors who do not respond to other treatment options. To determine the effectiveness of Miraculon-B compared to the standard of care, a late-phase clinical trial has been conducted, and now the data needs to be analyzed.

The goal is to assess whether Miraculon-B is more effective than the standard of care in shrinking solid tumors for patients who do not respond to other treatments. Additionally, the analysis aims to identify different patient sub-groups that may benefit more from the new drug. This knowledge will be crucial in formulating a comprehensive package to present to regulatory agencies and healthcare professionals, outlining the likely beneficiaries of Miraculon-B and proposing an effective strategy for its prescription.

To address this problem, two datasets were provided: clinical-study.csv and protein-levels.csv. The clinical-study.csv dataset contains information about patients' demographic characteristics, treatment group (Miraculon-B or standard of care), and response to treatment. The protein-levels.csv dataset provides data on the concentration of a potential predictive biomarker for solid tumors.

The analysis seeks to answer key questions:

1. Does the Miraculon-B treatment yield a superior treatment response compared to the control group?
2. Can patient factors such as age, weight, or protein concentration serve as predictors of treatment response?
3. Are there discernible patterns or sub-groups of patients that demonstrate a heightened response to Miraculon-B treatment?

By addressing these questions and extracting meaningful insights from the data, we aim to contribute to GSK's ongoing efforts to understand the efficacy of Miraculon-B. Furthermore, our analysis will aid in identifying patients who are most likely to benefit from this revolutionary treatment approach.

Approach:

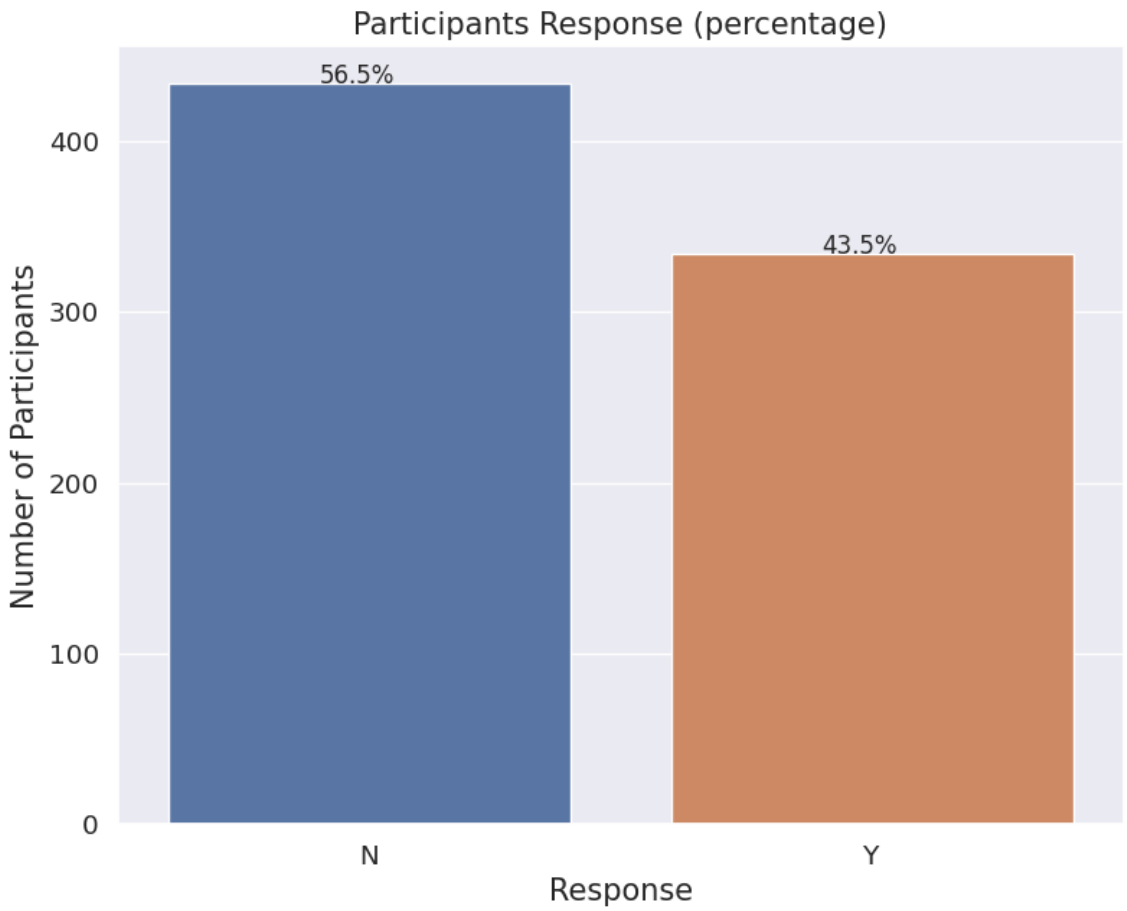
The analysis employed Python programming language within a Google Collaboratory notebook and Power BI, enabling independent data analysis. Advanced machine learning techniques were leveraged to develop a model capable of predicting the response of patients receiving Miraculon-B.

The analysis proceeded with data cleaning and pre-processing steps to ensure the integrity and reliability of the results. This included handling missing values, removing duplicate records to prevent data distortion, excluding participants below 18 years of age, and creating a new feature, 'Body Mass Index.' Once the data was cleaned, a structured approach was employed:

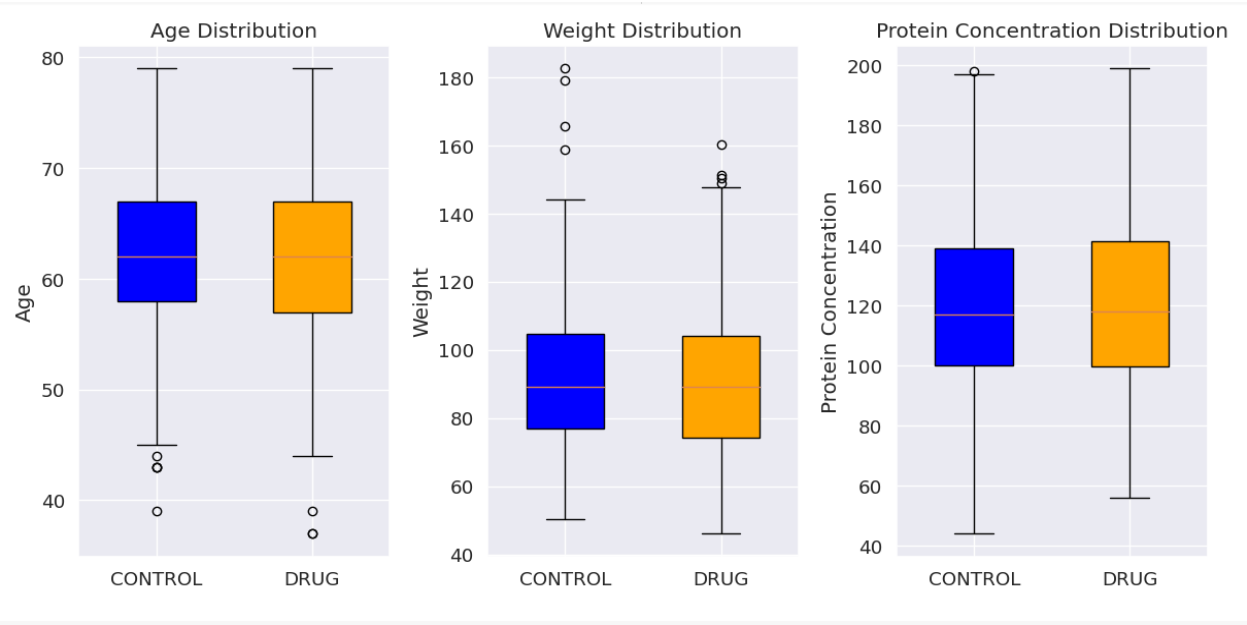
- Assessing the overall response of participants to treatment, irrespective of the drug used.
- Conducting preliminary analyses to eliminate potential biases by ensuring consistent characteristics across the treatment arms.
- Evaluating the response of participants within each treatment arm.
- Performing exploratory data analysis on participants treated with Miraculon-B.
- Conducting exploratory data analysis on participants receiving standard care.
- Identifying patterns and common behaviors among participants who responded to Miraculon-B treatment or standard care.

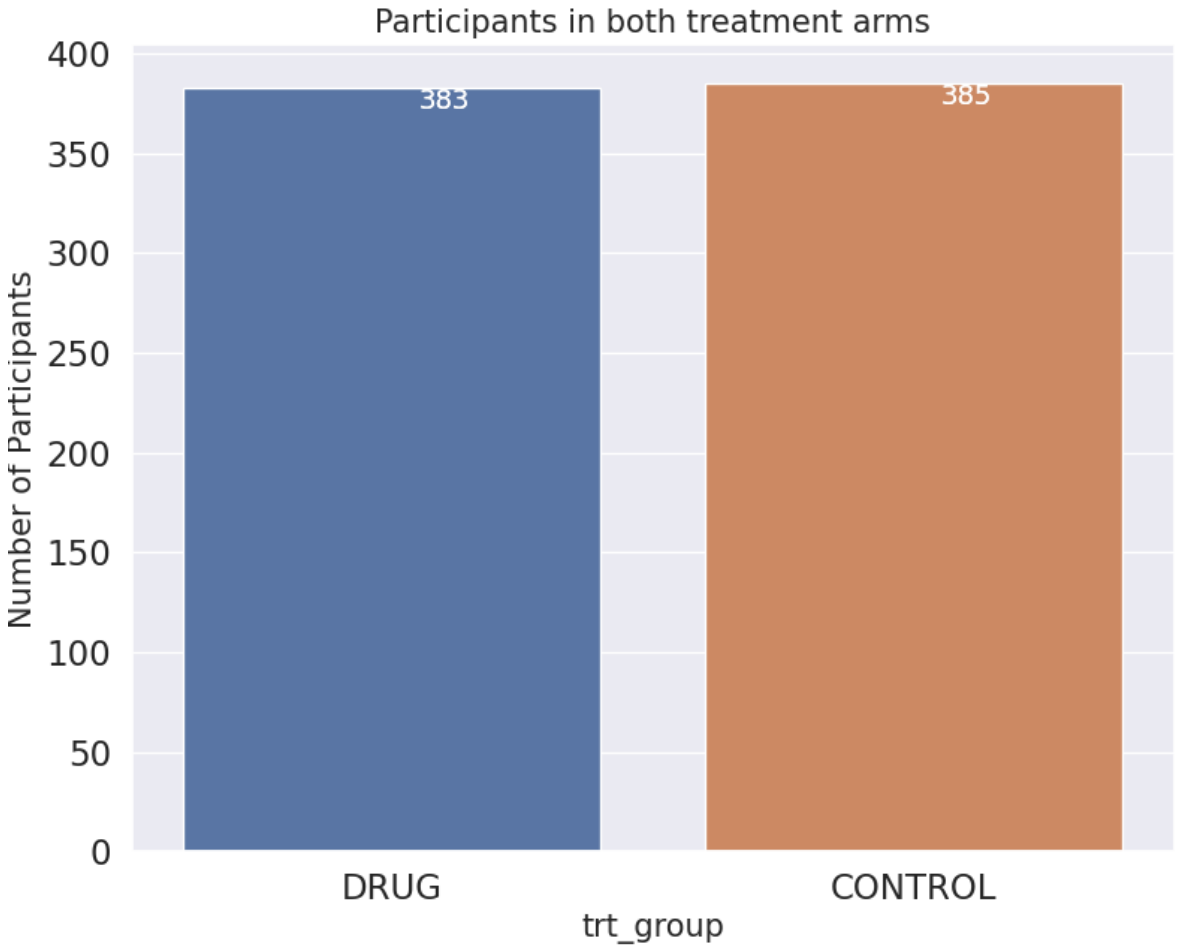
Data Analysis Findings:

1. Response of participants to treatment: 43.5% of participants demonstrated a positive treatment response.

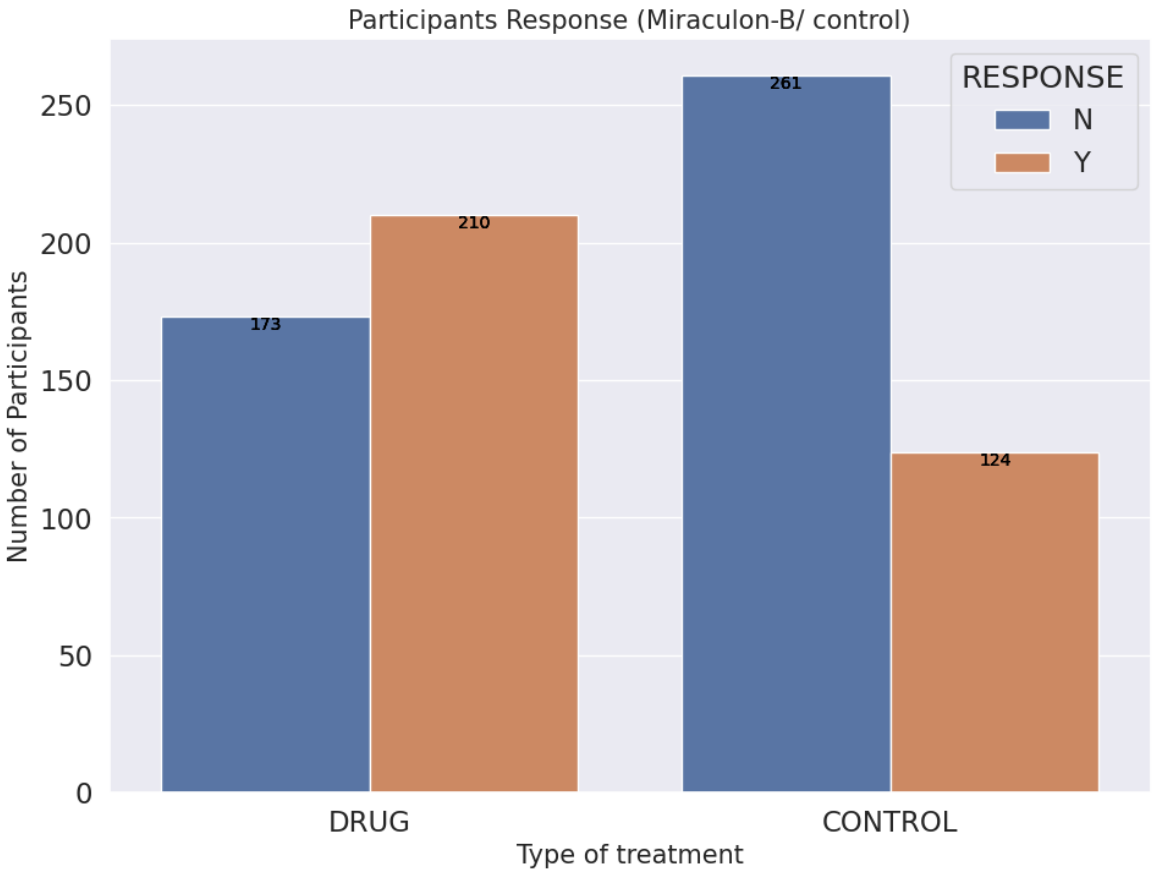


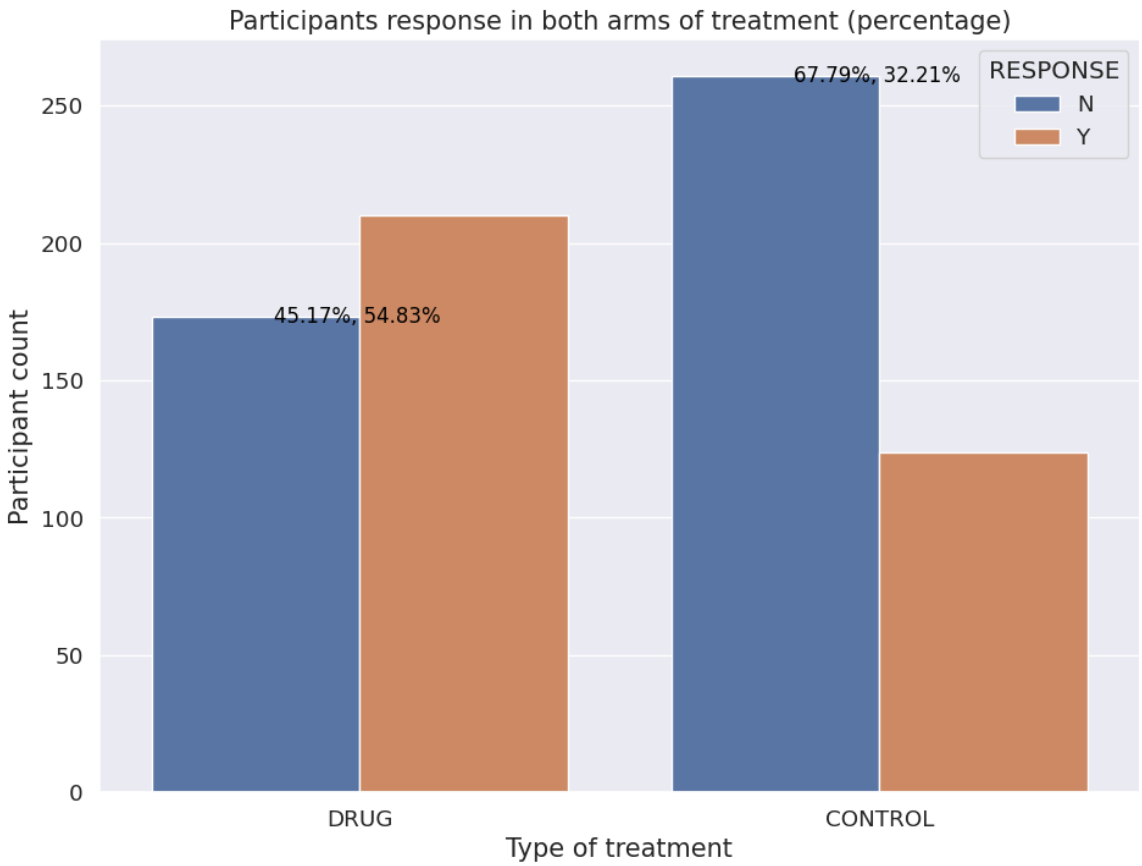
2. Preliminary Analysis: Boxplots revealed consistent mean age, mean weight, and mean protein concentration across both treatment arms. Count plots confirmed an equal number of participants in each treatment group.





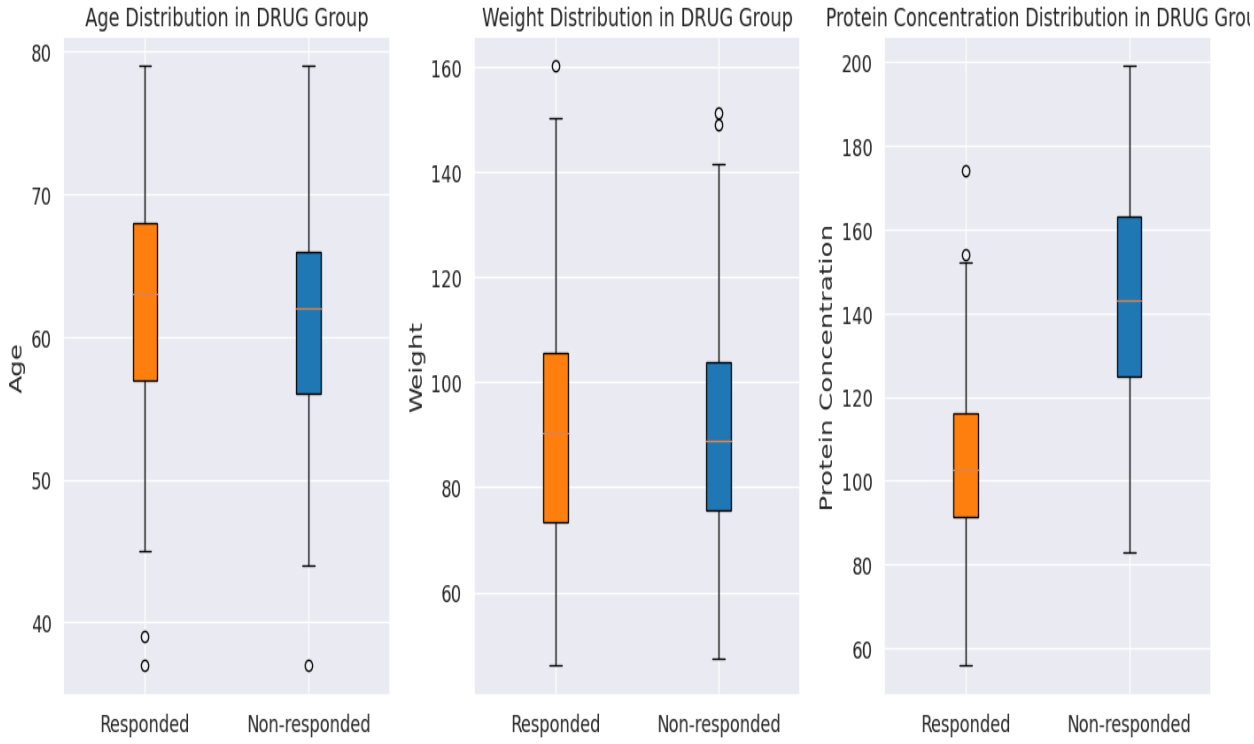
3. Response of participants by treatment arm: 54.8% of patients treated with Miraculon-B exhibited a positive response, while only 32.3% of those receiving standard care responded favorably.

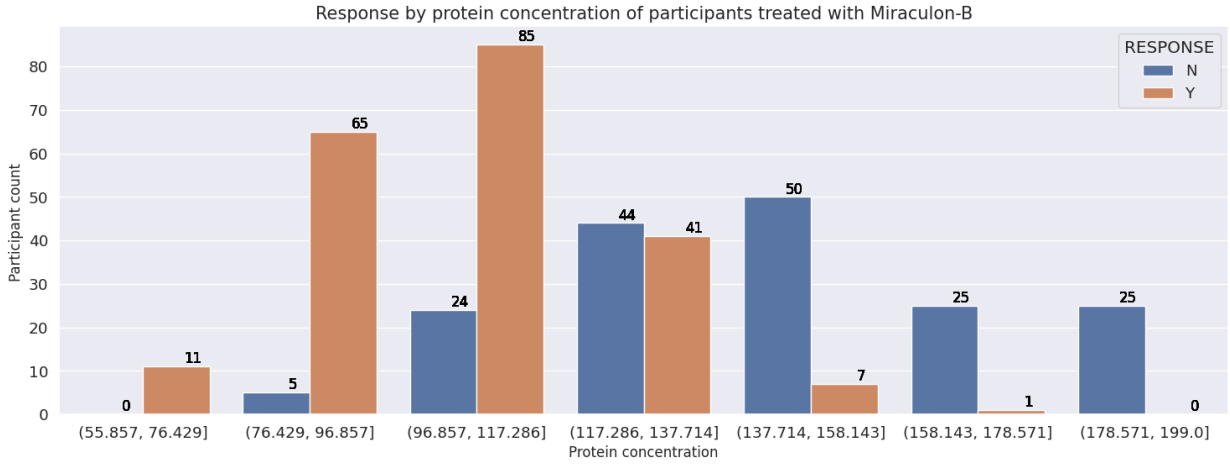

Participants Response (Miraculon-B/ control)
RESPONSE
N
Y
261
210
173
124
250
200
150
100
50
0
Number of Participants
DRUG
CONTROL
Type of treatment

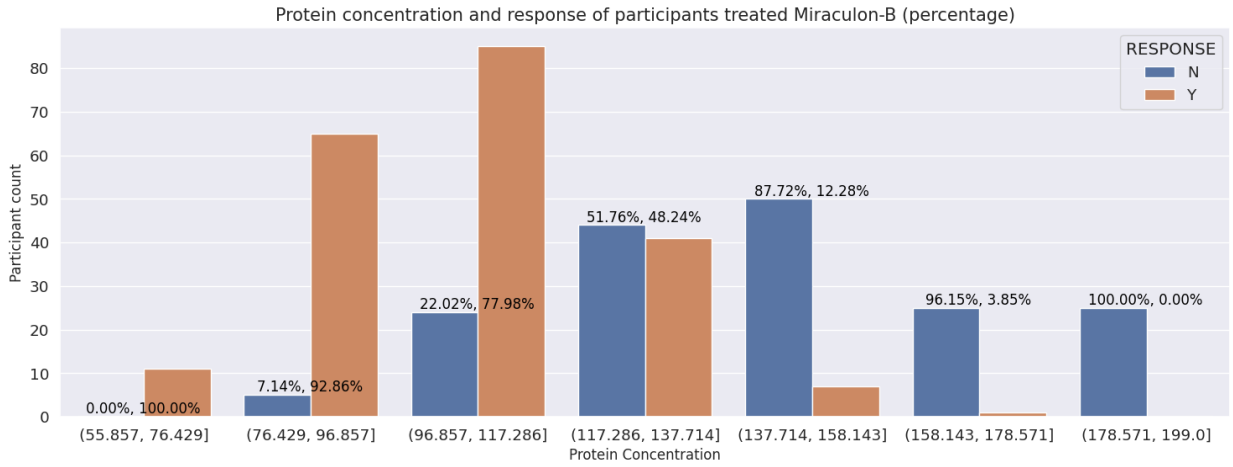


4. Response of patients treated with Miraculon-B: Responders and non-responders exhibited similar mean age and mean weight. However, responders displayed a significantly lower mean protein

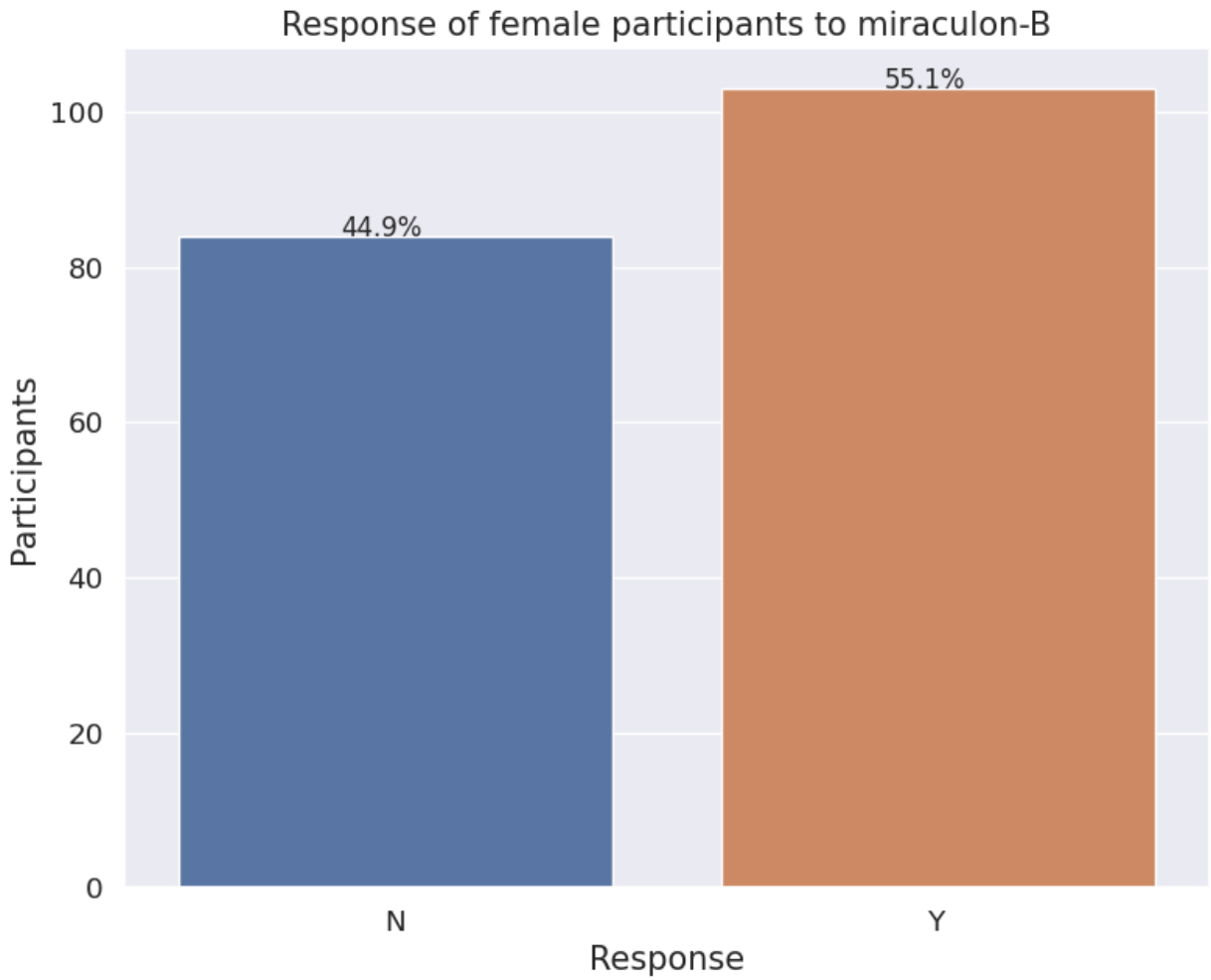
concentration (103.96 ug/L) compared to non-responders (144.07 ug/L).

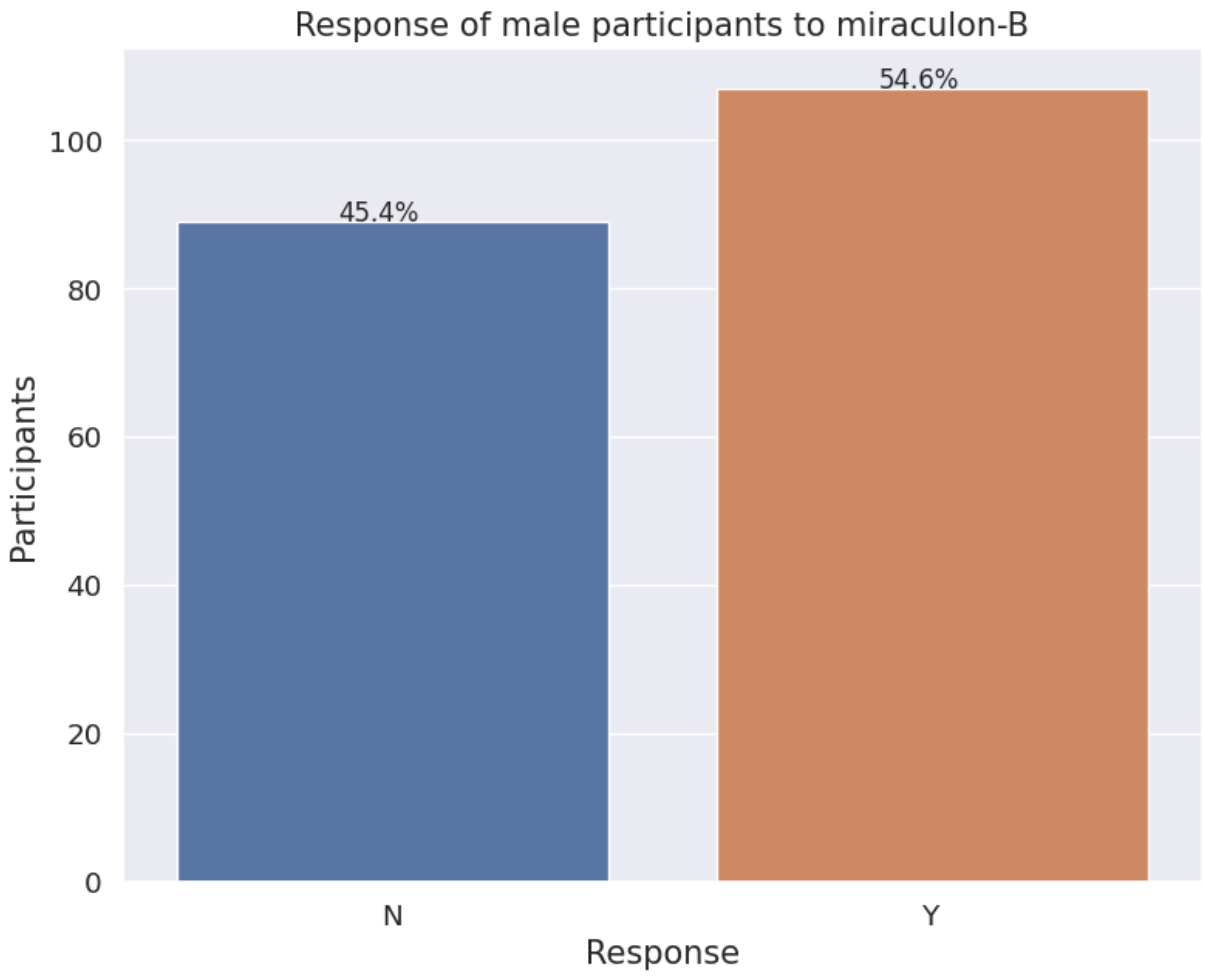




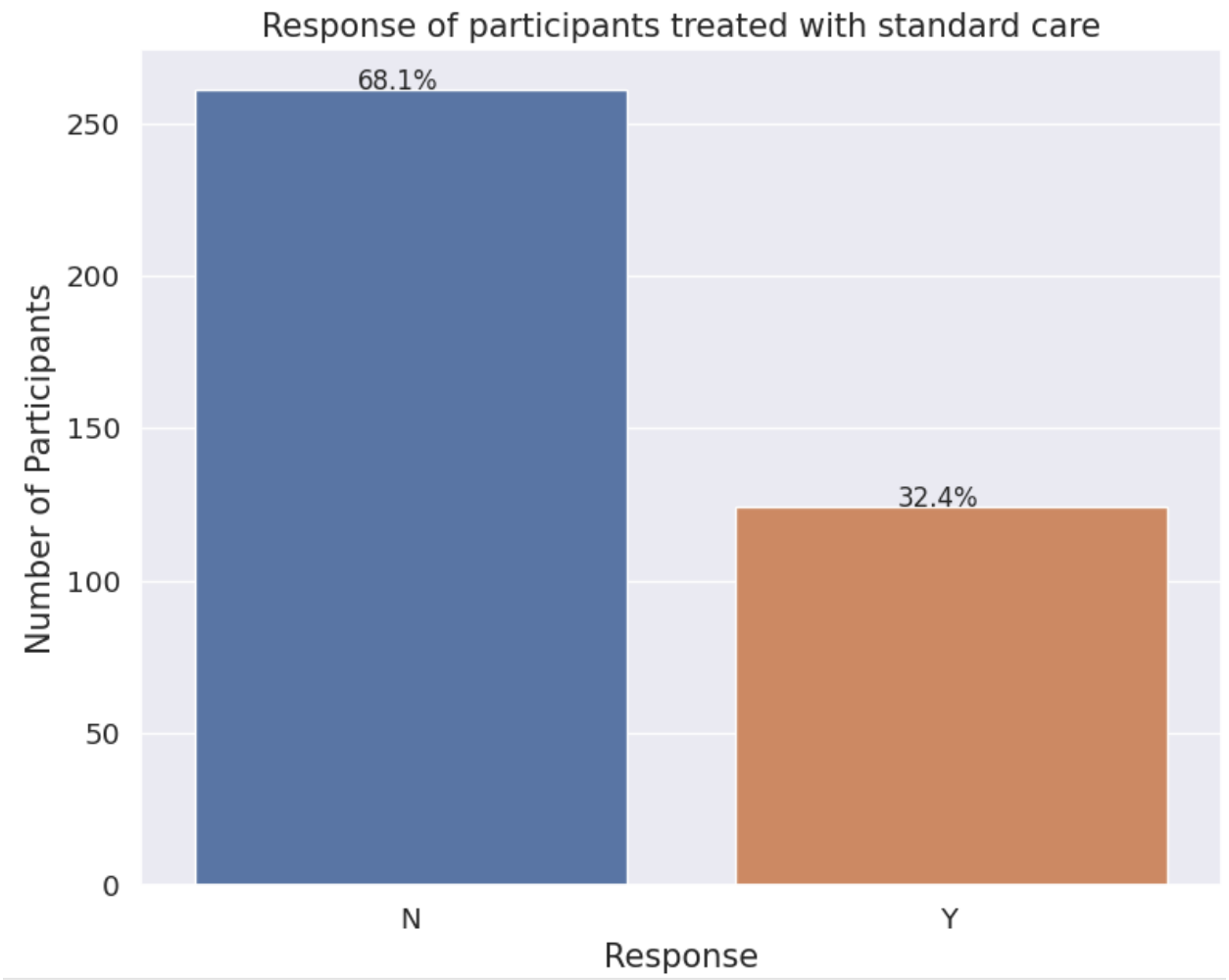


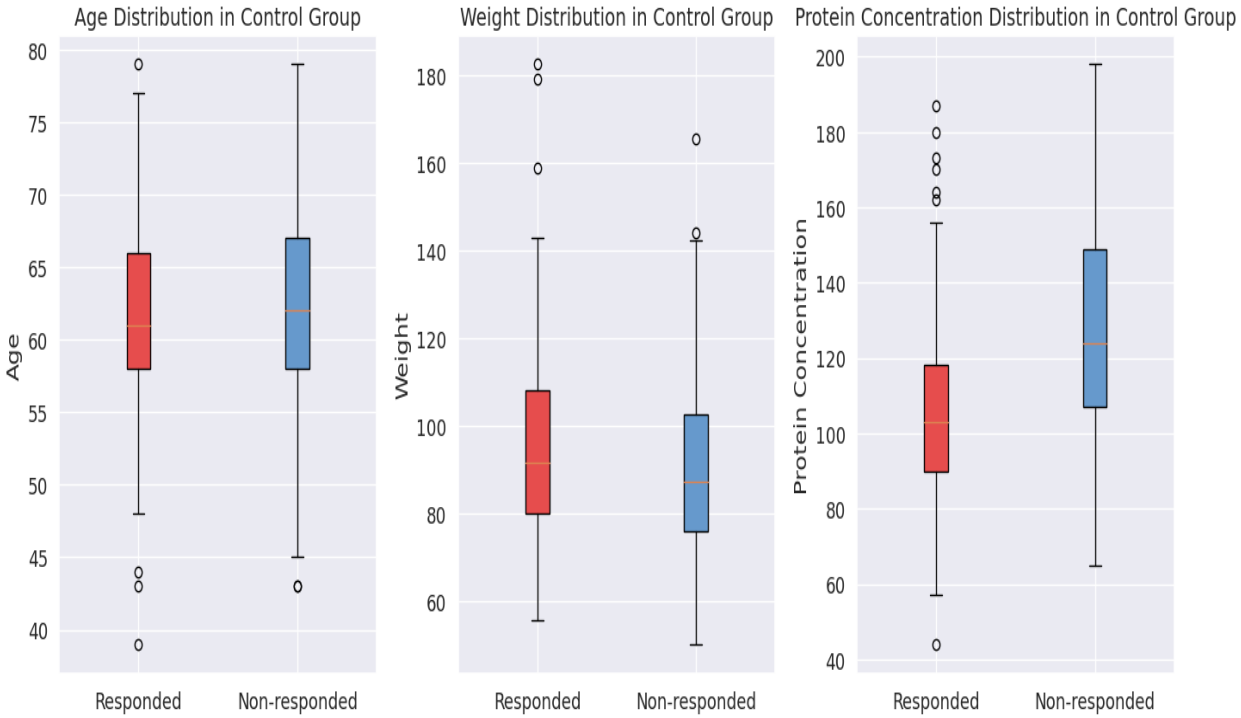
Percentage of male and female patients that responded to treatment is the same.

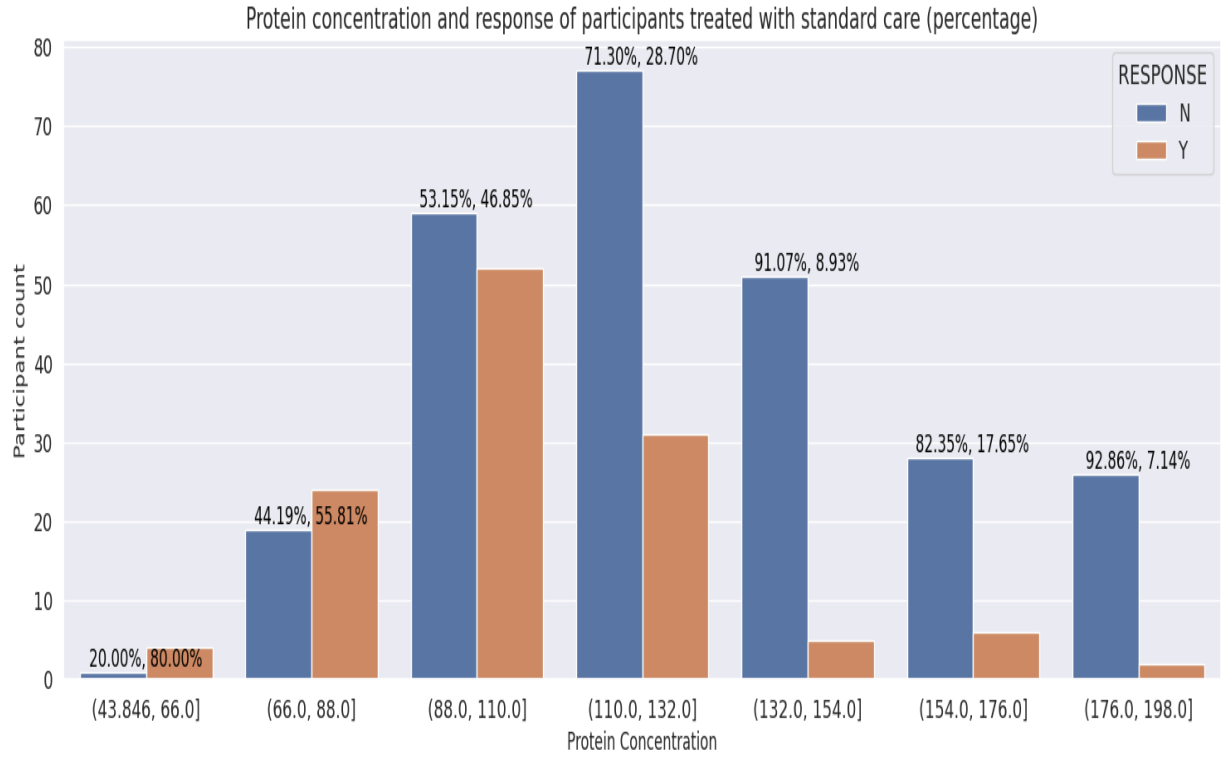

Response of female participants to miraculon-B
55.1%
44.9%
100
80
60
40
20
0
Participants
N
Y
Response



5. Response of patients treated with standard care:
Similar patterns and behaviors were observed among participants who responded to both Miraculon-B. Besides protein concentration, the mean age and mean weight for responders and non-responders is the same.


Response of participants treated with standard care
68.1%
32.4%
250
200
150
100
50
0
Number of Participants
N
Y
Response





Discussion of Results:

1. Miraculon-B demonstrated superior effectiveness in shrinking solid tumors compared to the standard care, with a remarkable 55% response rate compared to the control drug's 32% response rate.

2.Protein concentration can be used as a predictive biomarker. Lower protein concentrations (below 105 ug/L) were associated with improved treatment response in both treatment groups.

3. Among patients treated with Miraculon-B, those with a protein concentration below 96.87 ug/L exhibited the highest response rate: 100% positive response for concentrations between 55.857 ug/L and 76.43 ug/L, and a 92.8% response rate for concentrations below 96.857 ug/L.

4. Factors such as age, weight, and sex did not significantly impact treatment response.

Conclusion:

Analysis of the clinical trial data has uncovered the immense potential of Miraculon-B as a groundbreaking drug for the treatment of solid tumors. With a significantly higher response rate compared to standard care, Miraculon-B offers newfound hope for patients who have previously shown resistance to conventional therapies. The identification of protein concentration as a predictive biomarker opens up new avenues for personalized treatment approaches.

This study provides invaluable insights for clinicians and researchers, aiding in the identification of patient populations most likely to benefit from Miraculon-B.