**Finding Methods to Determine the Effect of Anti-Anxiety Medicine (Benzodiazepines) on Memory Recall**

**Motivation and Research Question**

Benzodiazepines are a class of drugs that are commonly prescribed to treat anxiety disorders, insomnia, and seizures. However, previous research has shown that side effects include impaired cognition, especially memory, and are prone to causing psychological and behavioral abnormalities (Breggin, 1998).

Thus, the research question asked, “How can we perform preliminary analysis of the effect of anti-anxiety medicine (benzodiazepines) on memory recall before testing it on real humans?” Since this question concerned human experiments, we started with a virtual dataset on Islanders that closely mimicked real-life humans. This project allowed us to practice analyzing the effect of this medicine without needing human consent so that in the future, we can extend these methods to a more complex design for the greater population.

**About the Dataset**

The data came from the “Memory Test on Drugged Islanders Data” dataset posted on Kaggle by Steve Ahn and collected by Mr. Almohalwas at UCLA (Ahn, 2019). Specifically, it came from “The Islands” virtual world created by Dr. Michael Bulmer at the University of Queensland, a virtual human population to practice statistics about epidemiology (CAUSEweb 2019; Bulmer).

It had a total of 198 rows and 9 columns, each row being an islander. The active treatment was islanders who took a drug in the benzodiazepine class: Alprazolam (Xanax, Long-term) or Triazolam (Halcion, Short-term). Alternatively, the control treatment was islanders who took a sugar tablet (placebo). Our potential outcome of interest was the difference in memory scores before and after drug intake.

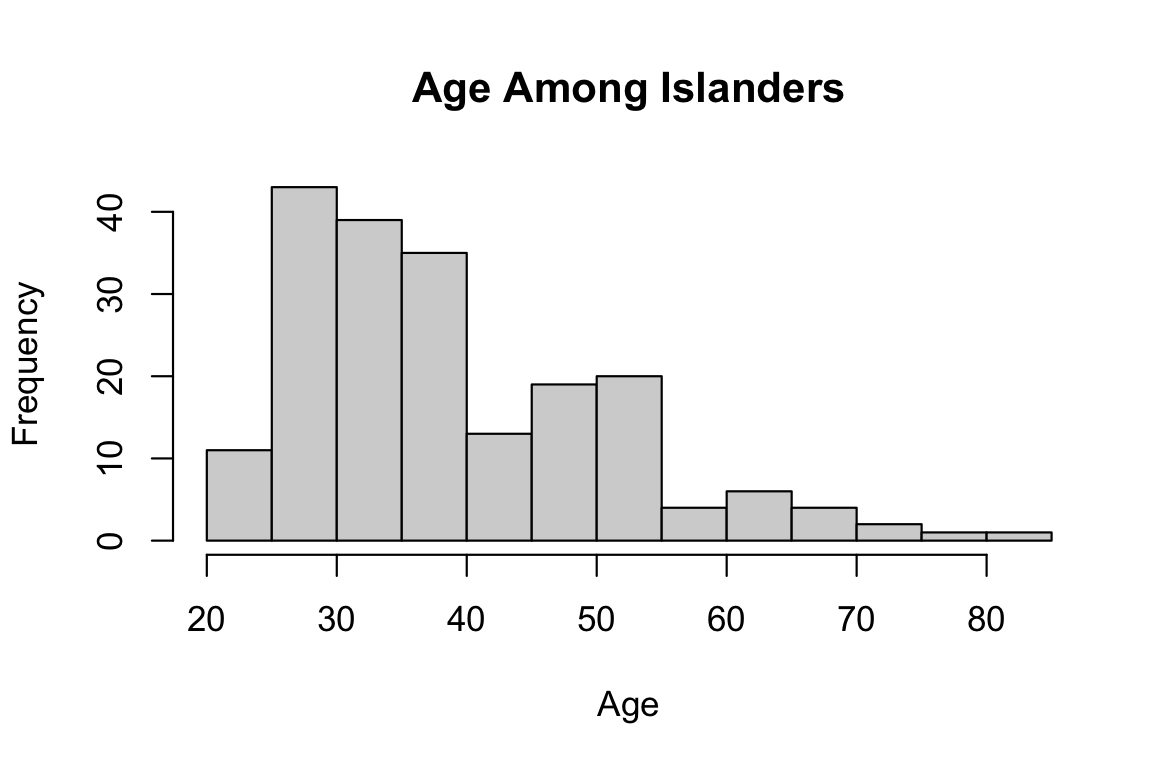
Other metrics that were measured included the islanders’ names, age, the type of memory they were primed 10 minutes before testing (happy or sad memories), the dosage of the drug (low, medium, or over recommended daily intake), memory score before and after the drug exposure. Both memory scores and its difference were measured by how many seconds it took to finish a memory test.

**Methods**

Step 1: Exploring and Cleaning the Data

Before performing analysis, we confirmed that the difference in memory scores should be the potential outcome of interest since we saw that there was a large correlation between memory scores before and after drug intake (r = 0.8).

We also noticed that the age variable was highly right skewed. Instead of only looking at mean age for covariate balance, we created a new variable called age groups. The histogram of ages seemed to show that they were distributed among three main groups generally well: 20-40, 40-60, and 60-83 (Figure 1). We wanted to make the age groups have the same range of years (approximately 20). The 60-83 group included one islander whose age was 83, the only person whose age was greater than 80. We thought it would be better to include them into this age group instead of making a new age group for them or throwing their data away.



**Fig 1.** Histogram of all ages of islanders.

Step 2: Determining Initial Covariate Balance

Since the data did not explicitly state that it was a randomized experiment and because we were unsure how the algorithm truly worked for the virtual world, we could not assume randomization. Thus, we first investigated the initial covariate balance between the active and control groups (Table 1).

| Covariate | Levels | Active | Control |
| --- | --- | --- | --- |
| Proportion of type of memory primed on | Happy | 0.5 | 0.5 |
| Sad | 0.5 | 0.5 |
| Proportion of type of dosage | Low (dosage = 1) | 0.34 | 0.33 |
| Medium (dosage = 2) | 0.33 | 0.33 |
| Over recommended daily intake (dosage = 3) | 0.33 | 0.33 |
| Proportion of age groups | Ages 20-40 | 0.62 | 0.70 |
| Ages 40-60 | 0.29 | 0.27 |
| Ages 60-83 | 0.10 | 0.03 |
| Mean age | | 40.58 | 37.42 |

**Table 1.** Initial covariate balance.

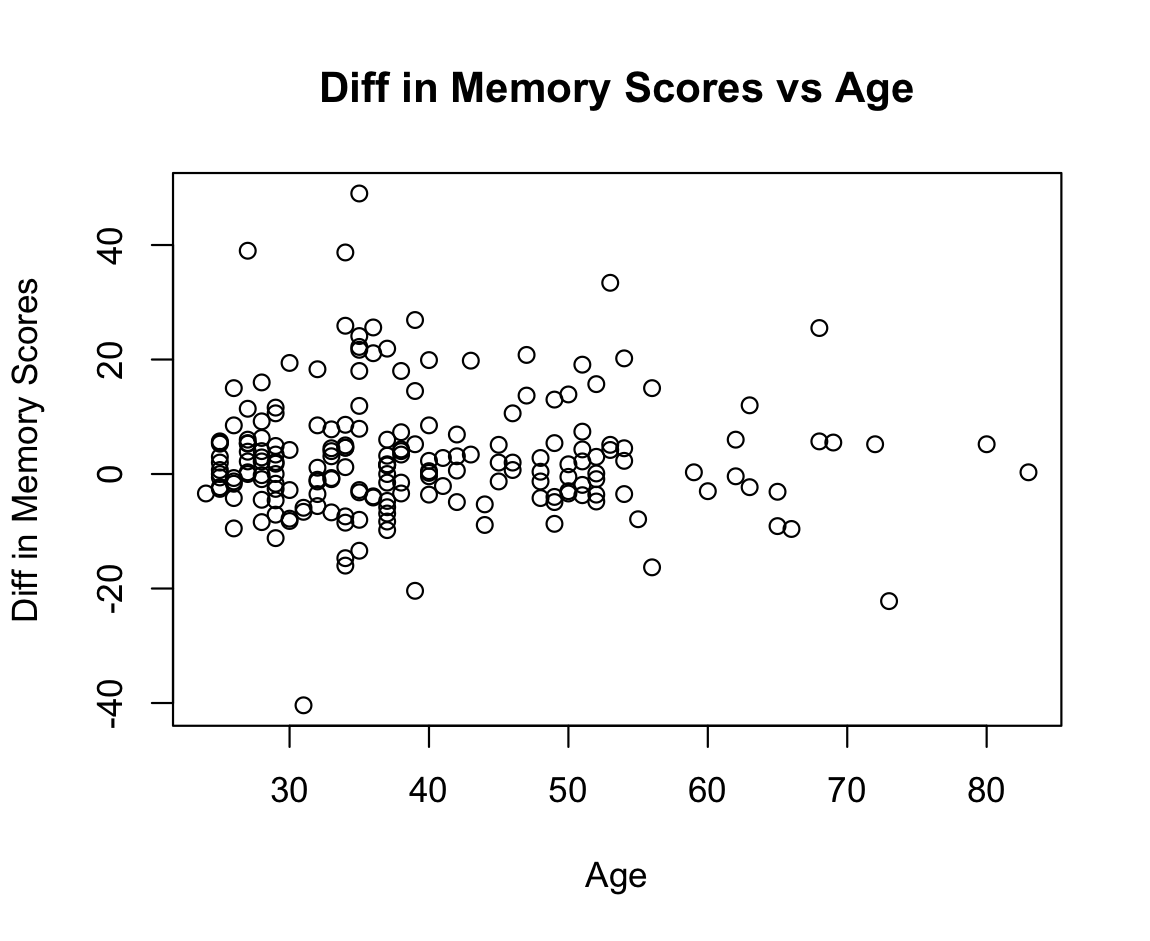
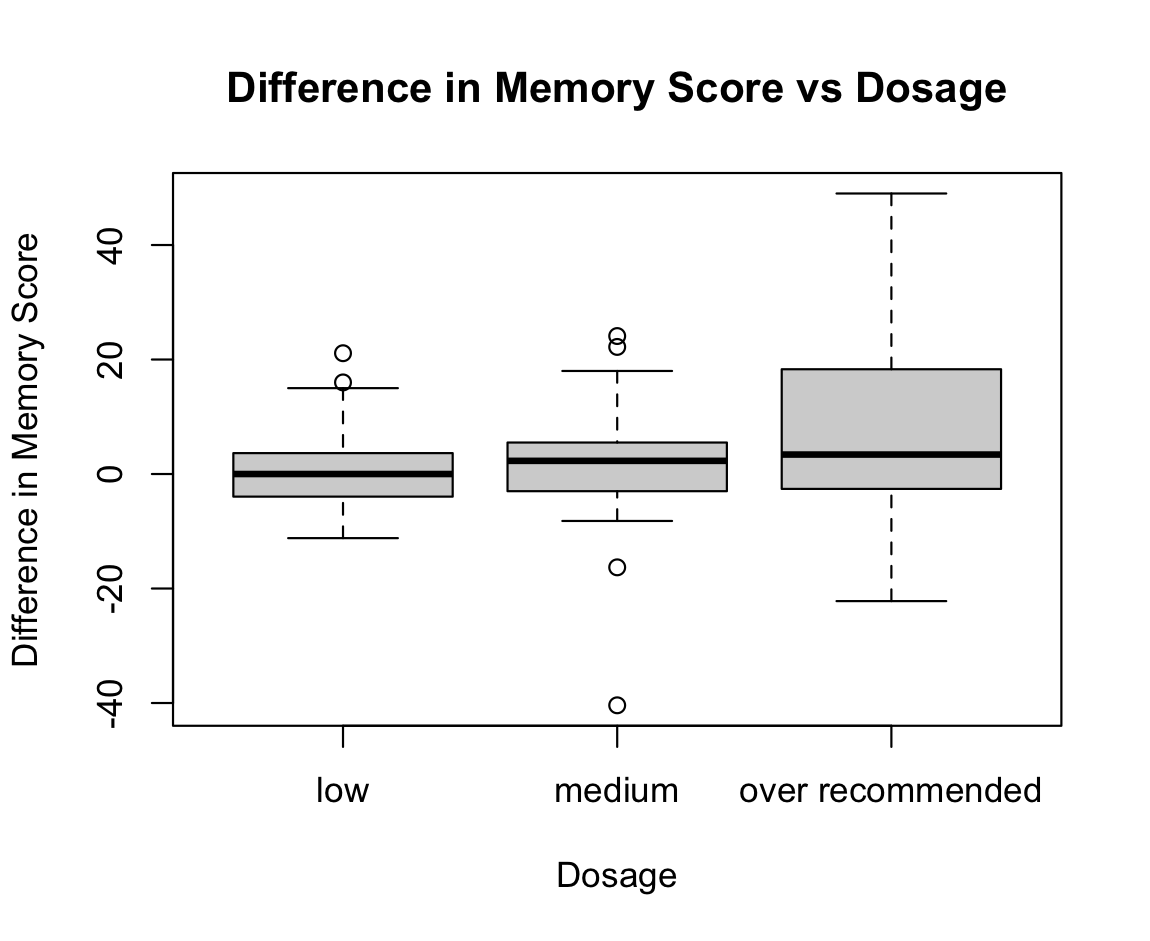
For the type of memory primed on and dosage, there was nearly perfect covariate balance. Although we did not have perfect covariate balance for age groups, subclassing did not satisfy better covariate balances. After trying many ways to estimate propensity scores and subclassing, the resulting covariate balances had worse balances. This was most likely because this was made-up data or was a randomized design that was stratified on these variables. This could also be due to the fact that there were only 66 control and 132 active units and that there were not a lot of covariates that were correlated with the potential outcome that we could choose to subclassify on.

Thus, we decided to keep the initial covariate balance and continue onto the next step.

Step 3: Multiple Imputation using Covariates

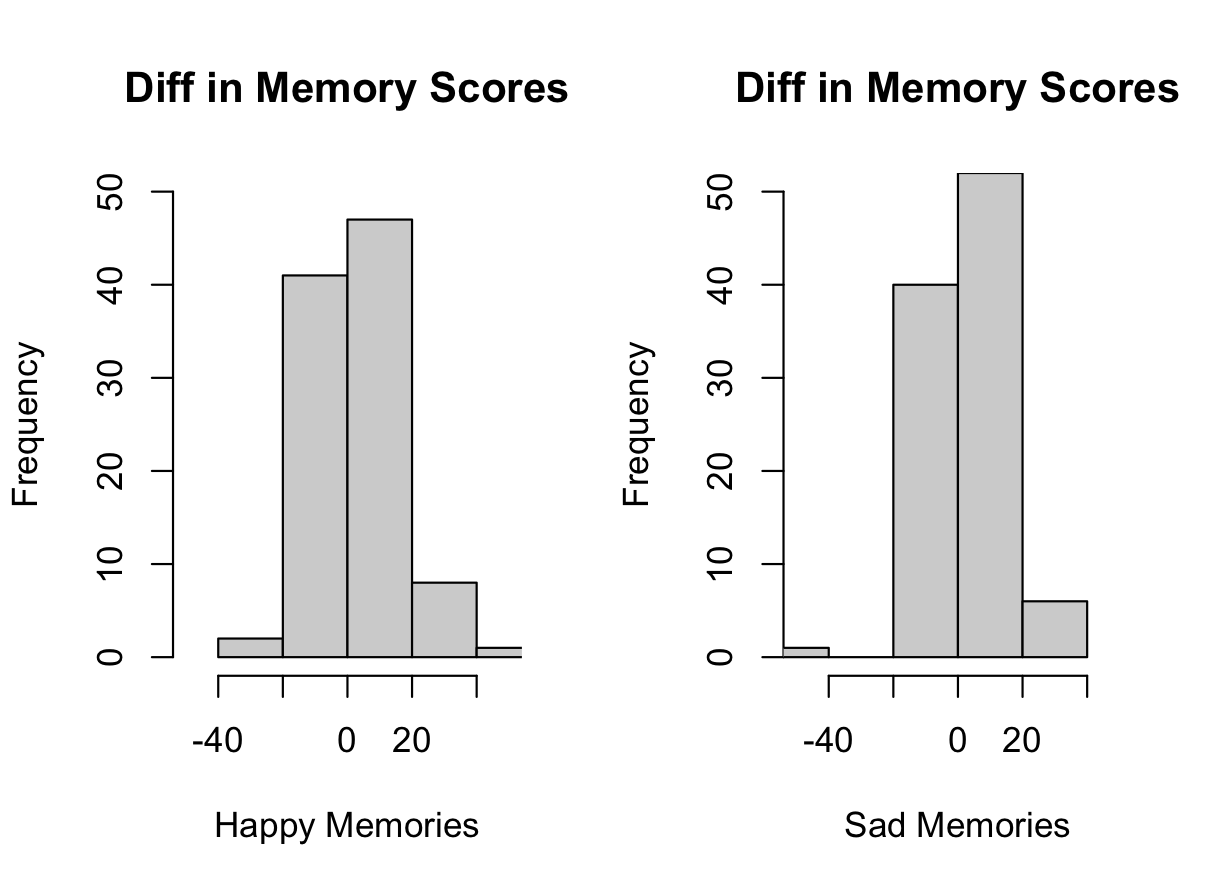
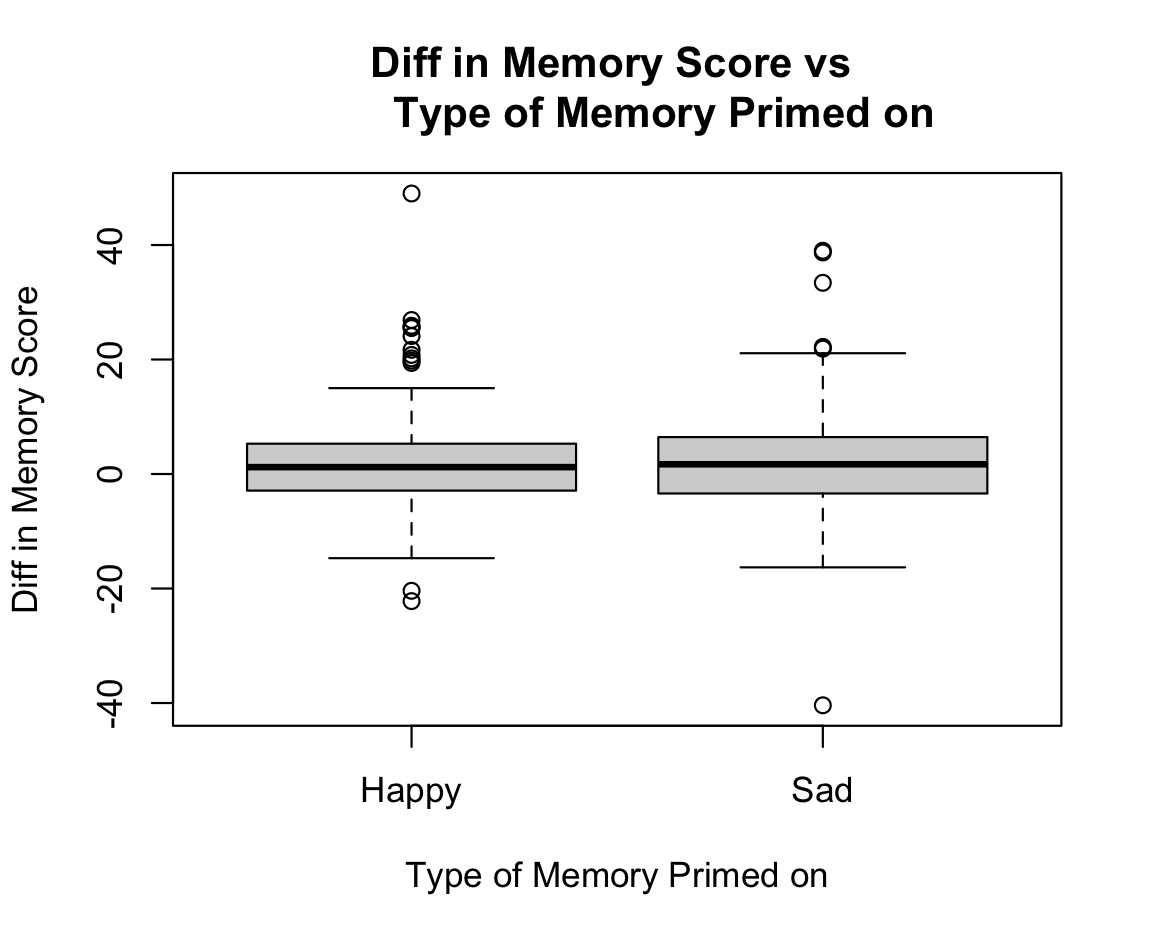
The initial covariate balance was very well balanced even though it could have been observational data, so we considered multiple imputation using covariates to obtain point estimates and 95% credible intervals.

We first looked at the correlation between the covariates and the potential outcomes. The boxplot showed that as the dosage increased, the difference in memory scores slightly increased (Figure 2). Age didn’t seem to be strongly correlated with the potential outcome (Figure 3), although the type of memory primed could have had a slight correlation based on the boxplot (Figure 4). Despite its weak correlation, we wanted to make use of the covariates because it was an observational study. We decided to use all three covariates (dosage, age, and type of memory primed) and compare the results from the three methods.



**Left: Fig 2.** Difference in memory score against dosage. As dosage increases, the difference in memory scores seems to slightly increase.

**Right: Fig 3.** Difference in memory score against age. There are no obvious signs of correlation between the covariate and potential outcome.



**Fig 4.** Difference in memory score against type of memory primed on using boxplot (left) and histograms (right). There seems to be a very slight increase in difference in memory scores for sad memories than happy memories.

Since dosage and type of memory primed are not continuous variables, we did not use separate regressions. In addition, because age was not linearly related to the outcome, we used age groups, a non-continuous variable, instead. We calculated the estimate of and 95% interval for the difference in means and medians using the empirical distribution with donor pools for the three observed covariates.

**Results**

| Covariate used | Estimate of diff in means | 95% interval for diff in means | Estimate of diff in medians | 95% interval for diff in medians |
| --- | --- | --- | --- | --- |
| Dosage | 4.70 | (3.52, 5.90) | 2.35 | (0.8, 3.7) |
| Age Groups | 4.81 | (3.56, 6.09) | 2.32 | (0.75, 3.70) |
| Type of memory | 4.70 | (3.48, 5.98) | 2.33 | (0.75, 3.70) |

**Table 2.** The calculated estimate of and the 95% interval for the difference in means and medians using dosage, age, and type of memory primed on as covariates.

From our preliminary analysis, the three covariates used (dosage, age groups, type of memory) were consistent with approximately 4.7 for the difference in means and 2.3 for the difference in medians. The 95% interval for the difference in means were approximately (3.5, 6) and (0.8, 3.7) for the difference in medians, respectively (Table 2).

**Conclusion**

To reiterate, the research question was to see how to perform preliminary analysis of the effect of anti-anxiety medicine (benzodiazepines) on memory recall before testing it on real humans. Memory scores were measured by how many seconds it took the islanders to finish a memory test. Thus, preliminary analysis showed that with the drug, people’s difference in memory scores before and after intake were 4.7 seconds more than people who received the placebo. This suggests that memory loss could be a side effect for benzodiazepines, but doesn’t seem to be a serious concern since four seconds is still a short amount of time. We also conclude that we have a general methodology of analyzing the effect of these drugs before testing it on real humans through using multiple imputation for an observational study.

**Discussion & Limitations**

A big limitation was that we did not know the assignment mechanism. It could have been randomized, but for an observational study, we had a lack of covariates and a small sample size.

We went through analysis assuming that dosage was a covariate, but we could not confirm this since we did not have enough information to know if it was determined prior to or after treatment assignment. As well, the multiple imputation method did not account for uncertainty about parameter estimates and restricted imputations to values observed in the dataset.

For the purpose of this project, assumptions were made before continuing onto the methods. However, not all assumptions might have held.

SUTVA may not hold, which is a crucial assumption to be able to use the causal framework. SUTVA I stated that for each islander, the drugs and placebo were administered the same way. However, we did not know the computer algorithm that administered the drugs or placebo, so we could not confirm this. SUTVA II stated that whether someone else gets a drug does not impact another islanders’ difference in memory score. However, based on their last name, there was a possibility that some islanders were family members – family members could talk to each other after getting a drug or not which could affect how well they remember certain memories, affecting their potential outcomes. On the other hand, we do not know how closely related the family members actually are in the virtual world and if they could even “talk” to each other, so we cannot completely accept this assumption as true.

Unconfoundedness also does not seem to hold because we do not know the assignment mechanism or the algorithm in the virtual world. We are missing the important covariates that would allow us to know the treatment assignment independent of the potential outcomes. Other possible covariates to measure include past medical history, prescription drugs currently taking, nutrition, or current cognitive health levels.

**Future Work**

In the future, we could extend these methods to a more complex design for the greater population that considers the limitations by recording more covariates and the exact assignment mechanism. Similarly, the issues for SUTVA mostly arise from the fact that this observational study was done in a virtual environment, so when done in real life, we could perform an experiment where we administer the drugs and placebo exactly the same and prevent interference between units as much as possible (e.g. restrict family members).

Hopefully, with these steps, we would be able to estimate the true average causal effect from an experimental design about how much benzodiazepines could impair cognition and memory loss as a side effect.

**References**

Ahn, Steve (2019). Memory Test on Drugged Islanders Data. Kaggle.

Breggin, P. R. (1998). Analysis of Adverse Behavioral Effects of Benzodiazepines With a Discussion on Drawing Scientific Conclusions from the FDA’s Spontaneous Reporting System. *The Journal of Mind and Behavior*, *19*(1), 21–49. http://www.jstor.org/stable/43854146

CAUSEweb. (2019). *Introductory Statistics Projects Using “The Islands” Virtual World Versus Student Collected Data*. *YouTube*. Retrieved from https://www.youtube.com/watch?v=cu75gsuMT3M.

Bulmer, M. (n.d.). The islands. https://islands.smp.uq.edu.au/login.php