Power Analysis W241 Project

Cynthia, Emanuel, Jonathan, and Rina

02/14/2024

1 Introduction

Our research question is: Does wearing sunglasses to limit visible light before bedtime improve sleep quality? We measure sleep quality via wearable activity trackers, such as a smart-watches and smart-rings. Sleep quality itself can be operationalized in various ways. We will be looking specifically at sleep quality as measured by:

1. Sleep Score: This is a measure from 0-100 calculated by common activity tracker platforms. The formulas are proprietary and can vary across platforms. However, we assume the methods used by each platform to be comparable as they measure the same features, such as heart rate, HRV, movement etc.
2. Proportion of sleep spent in REM and deep sleep stages: Time spent in specific stages of sleep, particularly REM and deep sleep, have been found to be essential for high quality sleep. (*Stages of sleep: What happens in a sleep cycle*)

Both sleep score and sleep stage proportions impose a notable risk of confounding sleep quality with varying activity tracker platforms and personal sleep habits.

1. Different platforms may calculate the same sleep conditions with different scores.
2. Individuals have varying schedules and sleep duration requirements. In addition, different sleep durations naturally affect sleep stage proportions. Deep sleep dominates earlier sleep, while REM dominates later. (*How much deep, light and REM sleep do you need?*)

We address this with a paired test, where each subject experiences both control and treatment phases. This allows us to increase power by controlling specifically for activity tracker platform and individual sleep habits.

1.1 Simulating Power

The significance level is set at 0.05. We consider the possibility that using sunglasses may also potentially harm sleep quality, thus we conduct two-tailed t-tests.

We simulate the experiment in different scenarios of true average treatment effect (ATE) and variance. For each scenario, we calculate power at different sample sizes, from 10 to 100 subjects. We run 1,000 simulations at each sample size. Note: Because we are running a paired test, each subject has a control and treatment value; for example, 10 subjects yields 10 control and 10 treatment sleep measurements.

2 Sleep Score Scenarios

Assumptions:

- The mean non-intervention sleep score is 75 (for scenarios 1-4). The standard deviation of sleep scores per individual is ~7.
  - These figures are determined by a gathering of 2-3 week sleep scores from friends of the researchers.
  - This average is also supported by data provided by Garmin (*Garmin Blog*).
- Sleep scores are roughly normally distributed in the population. It is bound by [0-100].

2.1 Scenario 1 (Base Case) - Measurement: Sleep Score, ATE: 8, SD: 7

This is the base case scenario.

- Mean non-intervention sleep score mean is 75
- True ATE = 8, based on a small trial run of our experiment conducted with ourselves.
- Standard deviation = 7, regardless of intervention

```
##     sample_sizes sample_powers
##  1:           10         0.518
##  2:           20         0.922
##  3:           30         0.996
##  4:           40         1.000
##  5:           50         1.000
##  6:           60         1.000
##  7:           70         1.000
##  8:           80         1.000
##  9:           90         1.000
## 10:          100         1.000
```

In this scenario, we are able to achieve a power of 0.922 with a sample size of 20.

2.2 Scenario 2 - Measurement: Sleep Score, ATE: 6, SD: 7

For this scenario, we test a case a more conservative ATE.

- Mean non-intervention sleep score is 75
- True ATE = 6
- Standard deviation = 7, regardless of intervention

```
##     sample_sizes sample_powers
##  1:           10         0.237
##  2:           20         0.591
##  3:           30         0.901
##  4:           40         0.961
##  5:           50         0.995
##  6:           60         1.000
##  7:           70         1.000
##  8:           80         1.000
##  9:           90         1.000
## 10:          100         1.000
```

In this scenario, we are able to achieve a power of 0.901 with a sample size of 30.

2.3 Scenario 3 - Measurement: Sleep Score, ATE: 4, SD: 7

For this scenario, we test an even more conservative ATE, half the base case.

- Mean non-intervention sleep score is 75
- True ATE = 4
- Standard deviation = 7, regardless of intervention

```
##     sample_sizes sample_powers
##  1:           10         0.070
##  2:           20         0.180
##  3:           30         0.367
##  4:           40         0.563
##  5:           50         0.718
##  6:           60         0.814
```

```
##  7:           70         0.905
##  8:           80         0.957
##  9:           90         0.974
## 10:          100         0.983
```

In this scenario, we are able to achieve a power of 0.814 with a sample size of 60.

2.4 Scenario 4 - Measurement: Sleep Score, ATE: 4, SD: 5

For this scenario, we take the previous scenario of lower ATE but assume a more forgiving variance, which potentially can be realized by not including weekend measurements, where sleep scores were consistently worse in trial data.

- Mean non-intervention sleep score is 75
- True ATE = 4
- Standard deviation = 5, regardless of intervention

```
##     sample_sizes sample_powers
##  1:           10         0.193
##  2:           20         0.530
##  3:           30         0.816
##  4:           40         0.928
##  5:           50         0.974
##  6:           60         0.996
##  7:           70         1.000
##  8:           80         1.000
##  9:           90         1.000
## 10:          100         1.000
```

In this scenario, we are able to achieve a power of 0.816 with a sample size of 30.

3 Deep & Rem Sleep Proportion Scenarios

3.1 Scenario 5: Measurement: Deep & REM sleep

We are interested in potentially looking at other sleep metrics as measurement of our outcome of sleep quality. We know that sleep quality, and not just duration, particularly the amount of time spent in the deep and REM phases of sleep can be strong indicators of sleep quality (*Suni, E.*). We attempted to used the effect size calculated by Hedges' g in the study by *Chinoy, E, et al.* in which they compared several different consumer sleep tracking devices to the laboratory standard of polysomnography.

For this scenario we use values from the *Chinoy, E, et al.* study

- The average non-intervention deep sleep phase as a percentage of total sleep is approximately 15.19%.
- The average non-intervention REM sleep phase as a percentage of total sleep is approximately 26.27%.
- The true ATE of deep sleep is 0.185 (minutes) or 0.04%.
- The true ATE of REM sleep is 0.25 (minutes) or 0.05%..
- Standard deviation of deep sleep = 6.02%, regardless of intervention.
- Standard deviation of REM sleep = 8.36%, regardless of intervention.

```
##     sample_sizes sample_powers
##  1:           10         0.002
##  2:           20         0.001
##  3:           30         0.003
##  4:           40         0.000
##  5:           50         0.000
##  6:           60         0.000
```

```
##  7:           70         0.002
##  8:           80         0.001
##  9:           90         0.000
## 10:          100         0.003
```

```
##     sample_sizes sample_powers
##  1:           10         0.003
##  2:           20         0.001
##  3:           30         0.005
##  4:           40         0.000
##  5:           50         0.000
##  6:           60         0.000
##  7:           70         0.000
##  8:           80         0.003
##  9:           90         0.000
## 10:          100         0.001
```

We had some difficulty in interpreting the units of the Hedges' g effect size and how it factors into our power calculation. From these power results, we can see we have extremely low statistical power. We also looked at the deep & REM phases of our self-conducted trial study and found that the difference in means between treatment and control for time spent in deep & REM phases of sleep was also extremely small (not even one-tenth of a percentage point). It is therefore unlikely we can feasibly use these metrics for our outcome.

3.2 Scenario 6 - Deep & REM sleep meaningful ATEs

In order to achieve power of around 60%-80% wot 20 to 30 people we would need the true ATE for each one of these to be:

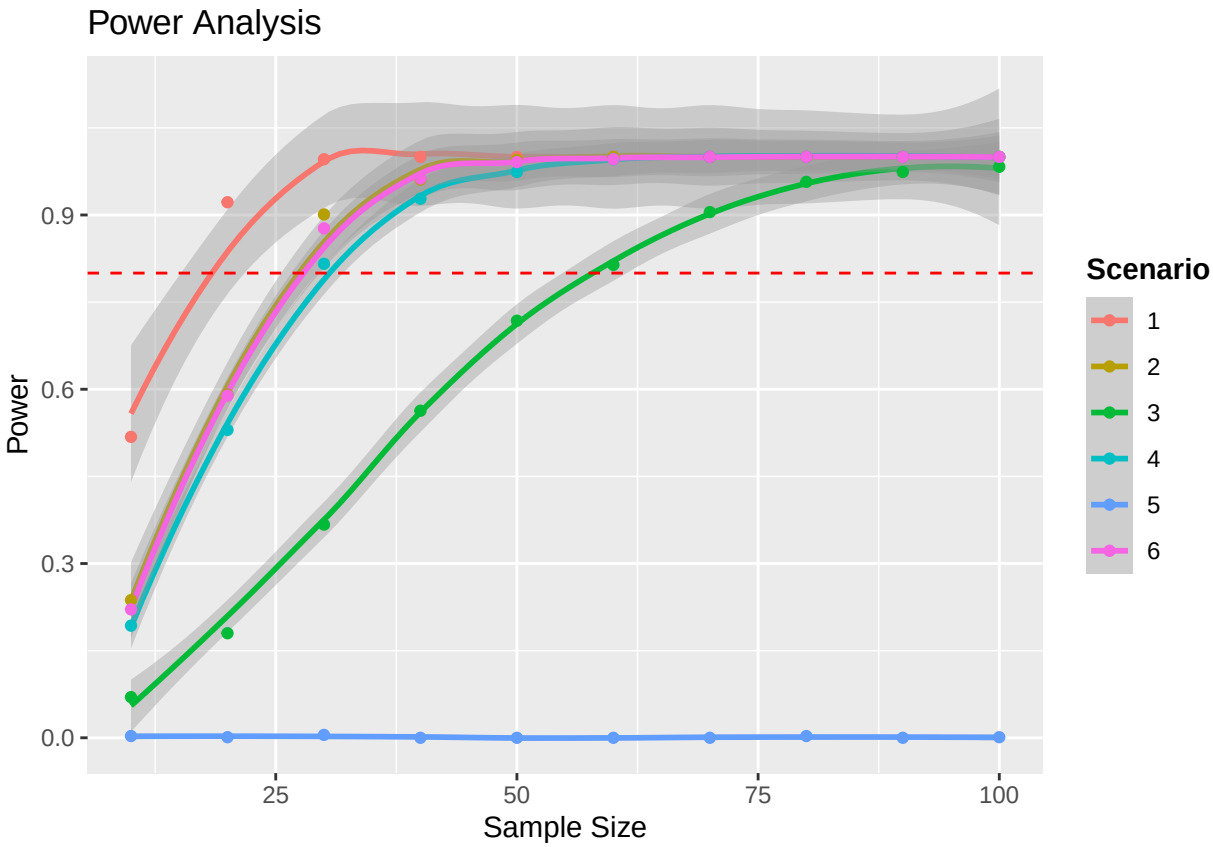
- $ATE_{deep} = 5\%$.
- $ATE_{REM} = 7\%$..

```
##     sample_sizes sample_powers
##  1:           10         0.202
##  2:           20         0.587
##  3:           30         0.852
##  4:           40         0.965
##  5:           50         0.989
##  6:           60         0.999
##  7:           70         1.000
##  8:           80         1.000
##  9:           90         1.000
## 10:          100         1.000
```

```
##     sample_sizes sample_powers
##  1:           10         0.221
##  2:           20         0.589
##  3:           30         0.877
##  4:           40         0.963
##  5:           50         0.991
##  6:           60         0.996
##  7:           70         1.000
##  8:           80         1.000
##  9:           90         1.000
## 10:          100         1.000
```

4 Plot of Achieved Power

Comparing the powers of each scenario, we plot:



We reach a power of .8 around 26-27 subjects.

5 References

Evan D Chinoy, Joseph A Cuellar, Kirbie E Huwa, Jason T Jameson, Catherine H Watson, Sara C Bessman, Dale A Hirsch, Adam D Cooper, Sean P A Drummond, Rachel R Markwald, Performance of seven consumer sleep-tracking devices compared with polysomnography, Sleep, Volume 44, Issue 5, May 2021, zsaa291, https://doi.org/10.1093/sleep/zsaa291

Miller DJ, Sargent C, Roach GD. A Validation of Six Wearable Devices for Estimating Sleep, Heart Rate and Heart Rate Variability in Healthy Adults. Sensors. 2022; 22(16):6317. https://doi.org/10.3390/s22166317

Suni, E., Singh, A. (2023, December 8). Stages of sleep: What happens in a sleep cycle. Sleep Foundation. https://www.sleepfoundation.org/stages-of-sleep

New data examines quality of Garmin users' sleep. Garmin Blog. (2023, August 30). https://www.garmin.com/en-US/blog/health/new-data-examines-quality-of-garmin-users-sleep/

Texas Health Resources. (2023, February 17). How much deep, light and REM sleep do you need?. Texas Health Resources. https://www.texashealth.org/areyouawellbeing/Health-and-Well-Being/How-Much-Deep-Light-and-REM-Sleep-Do-You-Need