You Snooze, You Lose. Right?

Estimating impact of snoozing alarms on individual's mental state

Anuj Ramakrishnan, Rohan Kar
Info 290 Experiments and Causality Final Project
Fall 2018, UC Berkeley

GitHub

Background

There are many factors that affect one's quality of sleep. The length of sleep, how tired one was before sleeping (physically and mentally), stress levels, noise levels, the quality of mattress/bed etc. are some of the factors that affect the quality of sleep. Yet we constantly optimize for the best possible sleep keeping in mind these factors while neglecting the effect of the snooze alarm we keep every morning. Is the humble snooze alarm ultimately responsible for a poor mental state after waking up? If so, why do we continue to keep snooze alarms? This has been the topic of great curiosity for many sleep experts and advocates who have spoken on platforms like the World Economic Forum and the Wall Street Journal asking to re-evaluate if the snooze alarm is doing more harm than good [1][2].

While we understand the obvious benefit alarms in helping us wake up on time, people continue to apply snoozing not fully understanding if it is actually helping their incomplete sleep or not.

Essential science behind this phenomenon suggests snoozing alarms break the sleep cycle or REM cycle. The extra sleep you get is not of the same quality which leaves one feeling more tired and "groggy" (this is called sleep inertia). Some studies have in fact showed self awakening can improve alertness and reducing sleep inertia [3].

Research on sleep inertia is important because it has also been crucial in detecting depression. Prior medical research has reported patients suffering depression have difficulty getting out of bed. Understanding a possible causal link between the snooze button and the effect on mental state may play an important role in helping diagnose a patient's depression symptoms.

Additionally, a study at the University of Surrey in the UK found that hitting the snooze button in the morning can affect cognitive functions throughout the day [4]. Rather than feeling more well-rested, you may have trouble concentrating or making decisions. For people having the need for alertness and active motor functions such as in doctors, or emergency services like police, firemen etc., it becomes critical to eliminate sleep inertia.

One key assumption involved in this experiment was setting the correct difference between alarms. This would effectively mimic a snooze button. The duration of this time gap is important as it defines how long before the next REM cycle is broken, hence affecting the how one feels after waking up. We had to ensure that the default of 9 minute snooze on iOS and 10 minutes on android were replaced with multiple alarms with 5 min differences between them.

Research Question

The research question can be summarized as:

Does snoozing alarms affect an individual's mental state after waking up as compared to not snoozing?

We hope to find a direct causal link between the effect of snooze alarms on the most critical aspect of sleep inertia: alertness (capturing acuity, attentiveness, motor dexterity etc.) and freshness (capturing fatigue and emotional state of well being). Prior sleep research has detailed the ideal physiological conditions for a good sleep. There have been numerous studies on environmental factors like the quality of mattress/pillow, air quality (humidity etc), noise levels and more but very few experiments have hypothesized snooze alarms on mental state.

Experiment

Experimental Design

The experiment follows a two group design, where the subjects are randomly assigned to either treatment (keeping 3 alarms / snoozing environment) or control (keeping just 1 alarm / no snoozing environment). The experiment was conducted over a period of 8 days and therefore data was collected from each subject for a maximum of 8 days (some subjects did not respond to the measurement survey on all days). The subjects were randomized 50:50 into treatment and control groups on a daily basis. Figure 1 below shows the experimental design using the ROXO grammar.

R	**O**	**X**	**O**
R	**O**	**-**	**O**

Figure 1: Experimental Design

In the sign-up survey sent out to prospective subjects, data was collected regarding covariates which could help explain the variance in the quality of sleep. Everyday, the subjects were sent an email informing them whether they had to keep **1 alarm** or **3 alarms** for that night. The following morning a short survey was sent out which had questions regarding the outcome measures (level of freshness and level of alertness), a question to check compliance and a few covariate questions.

An issue with this experiment was that the default snooze times were different in Android phones and iPhones. While the default snooze time could be changed on an Android phone (it could be set to any number between 1 and 30 minutes), it was not possible to change the snooze interval on an iPhone (it is fixed at 9 mins), At the same time, the snooze functionality could be switched off on an iPhone but it not possible to switch off the snooze functionality on an Android phone. To ensure the same experience for all subjects, it was decided that the snooze functionality will be mimicked by keeping multiple alarms at equal intervals of time. Subjects with an iPhone were instructed to switch off the snooze functionality while those with an Android phone were instructed to keep the snooze time as 30 minutes (this was an assumption in the experimental design that a 30 minute snooze time would be as good as not having a snooze function).

There were certain risks to the validity of the experiment.

First, it was necessary that all subjects fill in the response to the outcome question as soon as they wake up. Having a coffee or having a bath before filling out the survey might affect the response quality as the fatigue level could reduce because of undertaking the aforementioned actions. With participants in over 9 timezones, it had to be ensured that the daily survey was automated to be sent at the same time (locally) for every participant. As sleep inertia can last typically between 30 min up to a few hours, it was decided that 6 AM local time was the best time to send daily survey emails. The outcome of tiredness (sleep inertia) was captured by recording how an individual felt in terms of alertness and freshness on a 5 point likert scale just after waking up. There were cases where subjects missed sending a response on a day and sent two responses the next day (in the hope of responding for the missed day). In such cases only the last response from that particular day was saved and used in the analysis (and the other responses were discarded). This was done to ensure that the data being analyzed is valid and as close to reality as possible.

Experimental Materials

The experiment required the subjects to have a phone (to set the alarms) and access to emails and internet to receive the treatment assignment emails and the daily survey emails.

In this section, we will provide information regarding the randomization procedure, treatment and control assignment and the daily survey emails.

Randomization Procedure

All the subjects' names present at the end of the day were copied into a spreadsheet. 1 or 0 was assigned to each name at random by using the `np.random.shuffle( )` function in the Numpy package. The figure below shows the piece of code used to randomize treatment assignment.

```
x = np.repeat([0,1], repeats= [29,29])
x

array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1])

np.random.shuffle(x)
x

array([0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 1, 1, 0, 0, 0, 1, 0, 1, 1, 1,
       1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1, 1, 0,
       1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1])
```

Figure 2: Python code for randomizing treatment assignment

Randomizations were done daily with a 50-50 split of treatment and control. Treatment assignment emails were then sent out to the two groups. Figure 3 shows the daily routine of treatment assignment and measurement collection. The randomization was tweaked on the weekend to ensure that the subjects who were in treatment on Saturday were assigned to be in control on Sunday and vice-versa. This was done to ensure that all subjects were in treatment and control, both, over the weekend.

Weekend*

Date		11/28	11/29	11/30	12/01	12/02	12/03	12/04	12/05	12/06
R	Morning		O	O	O	O	O	O	O	O
	Evening	X	X	X	X	X	X	X	X	

Figure 3: Schedule of treatment assignment and measurement collection
**The treatment assignments were flipped over the weekend*

Treatment Emails

Figure 4 shows the treatment and control emails used in this experiment. The mails contained information regarding the number of alarms that have to be set for the given day and other instructions which ensure that all participants have as similar an environment as possible. This includes setting the volume of phone alarm to maximum and steps to change the snooze alarm setting for Android phones and iPhones. The mail also urged the participants to respond to the daily survey email as soon as possible after walking up. The messaging in the two emails was exactly the same except the line regarding the number of alarms to set.

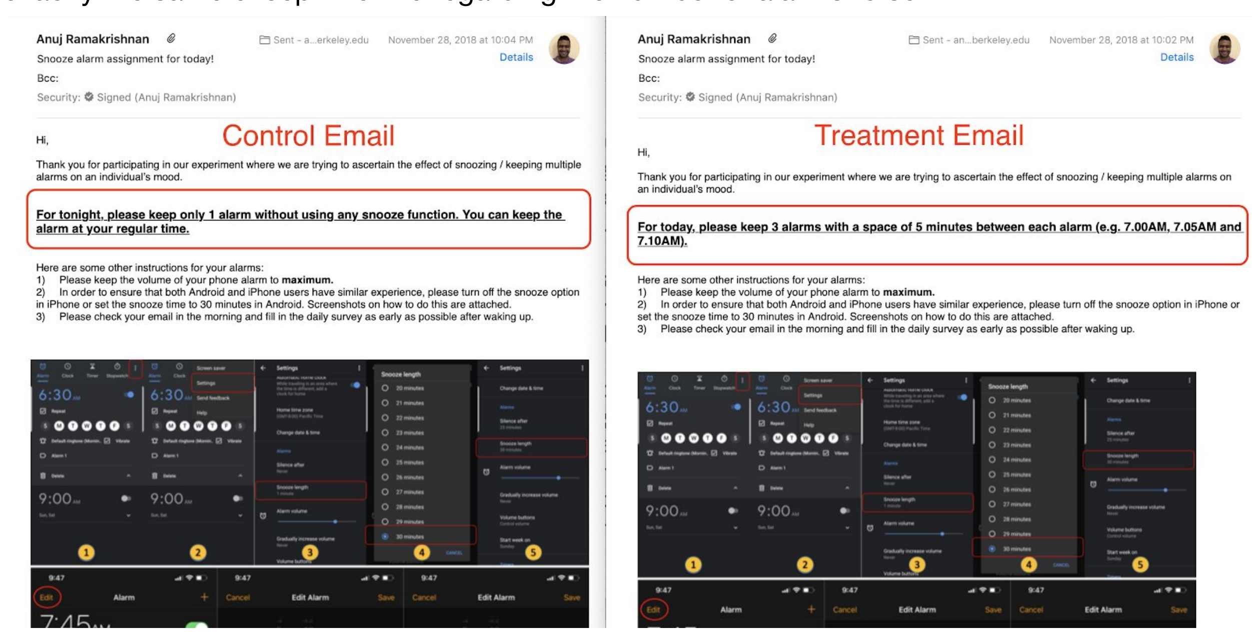

Figure 4: Screenshots of the control and treatment emails

- Subjects in treatment were asked to keep 3 alarms with a difference of 5 mins (to simulate a snoozing environment)
- Subjects in control were asked to keep just one alarm (to simulate an environment where they cannot snooze)

Daily Survey Emails

Everyday morning at 6AM, a short survey was sent out to all the participants which contained questions which measured the outcomes of interest (freshness and alertness), besides a few other covariates. Table 1 shows the set of questions which were asked in the daily survey email.

Question	Response Type	Options
How many alarms did it take to wake up?	Multiple Choice	None/1/2/3/4 or More
How refreshed do you feel right now?	5 point Likert	Very tired to Very refreshed
How alert do you feel right now?	5 point Likert	Not at all to Very Alert
Do you do any strenuous activity yesterday?	Binary	Yes/No
How long did you sleep?	Text box	Hh:mm
How stressed were you last night?	5 point likert	Very stressed to not at all stressed

Table 1: Questions from the Daily Survey Email

Q2 and Q3 in the table above are the primary outcome measures for our experiment. Q4, Q5 and Q6 were asked to be used as covariates.

**The complete list of covariates which were collected during the pre-treatment survey and daily survey is present in the appendix*

Compliance Check

Question 1 in Table 1 above was asked to serve as a compliance check. The overall aim of the experiment is to gauge the effect of snoozing alarms and hence we defined an answer of 'None' or '1' to question 1 as 'Did not snooze' while an answer of '2', '3' or '4 or more' was defined as 'Snoozed'.

For our experiment, compliers would be defined as those subjects who 'Snoozed' when assigned to the treatment group and 'Did not snooze' when assigned to the control group.

Learnings from Pilot

We ran a small pilot test with our initial sign-ups for the experiment. The feedback we received on the pre-treatment (sign-up) survey and treatment assignments were helpful and were incorporated. We realized that none of the participants would be ready to comply to an experiment where the control group had to keep 0 alarms. This was our ideal scenario since it would ensure that the control group has no chance of snoozing (assuming someone had not turned off the snooze functionality). We then decided to have the control group keep one alarm and mention in the mails that the snooze functionality should be disabled. We also received feedback on some covariates we should be measuring and added them to the pre-treatment survey.

Tracking Observations over the course of the Experiment

Our initial goal was to only include current students of UC Berkeley. We felt that this population represented similar work schedules, stress levels and on a feasibility side were likely to agree to participate and continue to respond to our experiment instructions with minimum churn.

However, for our final selection we decided to expand to a much broader population. This was done mostly to increase our sample size but we also felt that getting a more diverse set of participants would help with the generalization of our results. Our final sample size consisted of 59 participants signing up who had ages between 18 - 68 and had professions like Students, Professionals, Home-makers, Teachers etc. Figure 5 shows the flow of observations and measurements.

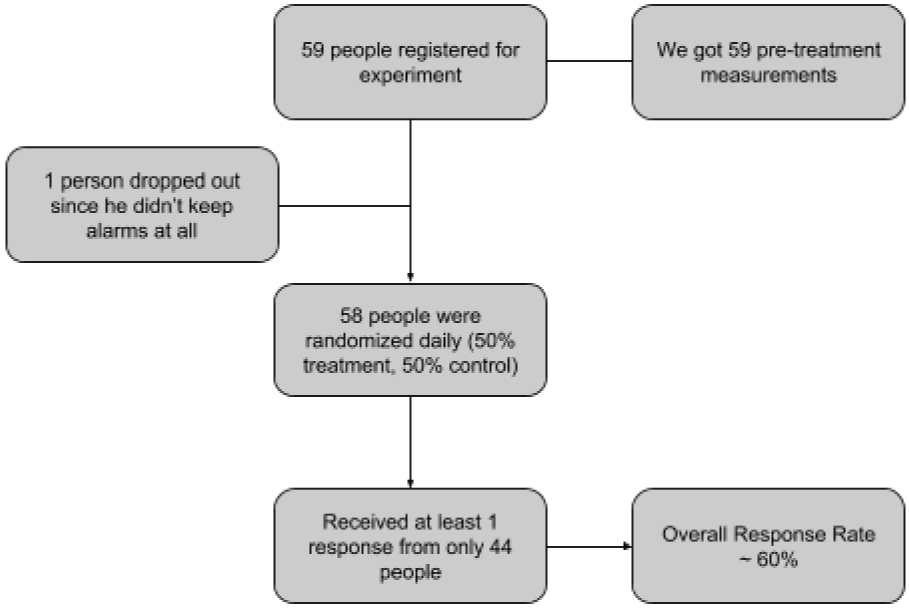

Figure 5: Tracking observations in the experiment

Results

Covariate Balance Check

We ran the randomizations at a daily level across all the participants. A covariate balance check helps to see whether the randomization was successful or not. An imbalance, if present can be corrected during the model building stage by controlling for imbalanced covariates. Table 2 shows the mean values of the numerical variables present in the dataset.

	age		avg_sleep		cat_deep_sleeper	
treat	**0**	**1**	**0**	**1**	**0**	**1**
date						
11/29/18	31.23	36.92	6.99	7.00	0.71	0.62
11/30/18	38.84	31.87	6.88	7.16	0.63	0.81
12/1/18	35.08	34.59	7.04	6.83	0.92	0.64
12/2/18	34.09	39.09	6.90	7.04	0.65	0.83
12/3/18	36.19	39.50	6.86	6.94	0.69	0.71
12/4/18	34.00	38.65	7.15	6.66	0.82	0.53
12/5/18	34.29	40.47	7.29	6.72	0.64	0.67
12/6/18	33.17	32.31	6.92	7.21	0.58	0.75

Table 2: Mean values of covariates

Visualizing Outcome Measures

Our two main outcome measure were a) a measure of freshness and b) a measure of alertness. The distributions of these outcome measures across treatment and control groups is shown below.

For the refresh outcome measure:

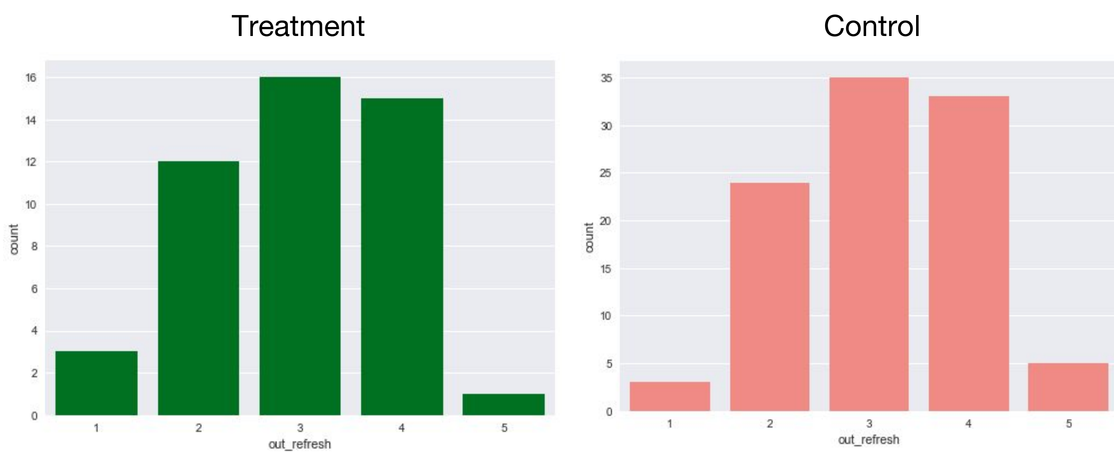

Figure 6: Distribution of the variable out_refresh amongst compliers

For the alert outcome measure:

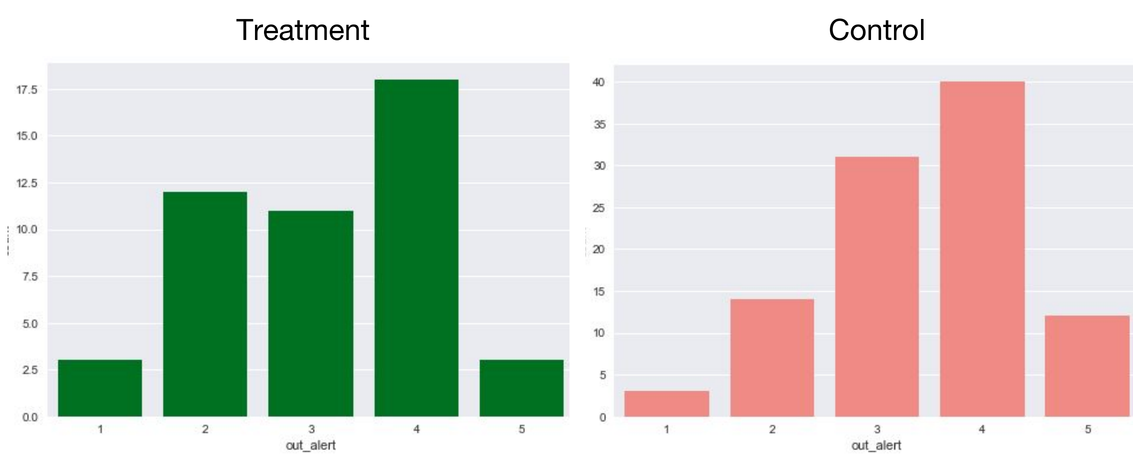

Figure 7: Distribution of the variable out_alert amongst compliers

From the above distributions of outcome measures, it seems more likely that we will see an effect on the alertness measure than the freshness measure.

Compliance

Our experiment brings forward a unique type of two sided non-compliance. In the case of our experiment, even if a subject in the treatment group keeps the required number of alarms, there is still the possibility that he/she wakes up before the alarm or with just one alarm. This subject would hence become a non-complier even though they had taken all the steps necessary to comply. Similarly, there might be subjects in the control group who do not wake up with their one alarm and might wake up due to some other stimulus. On analyzing the compliance across treatment and control groups, we see that we are pretty limited in our intent to treat.

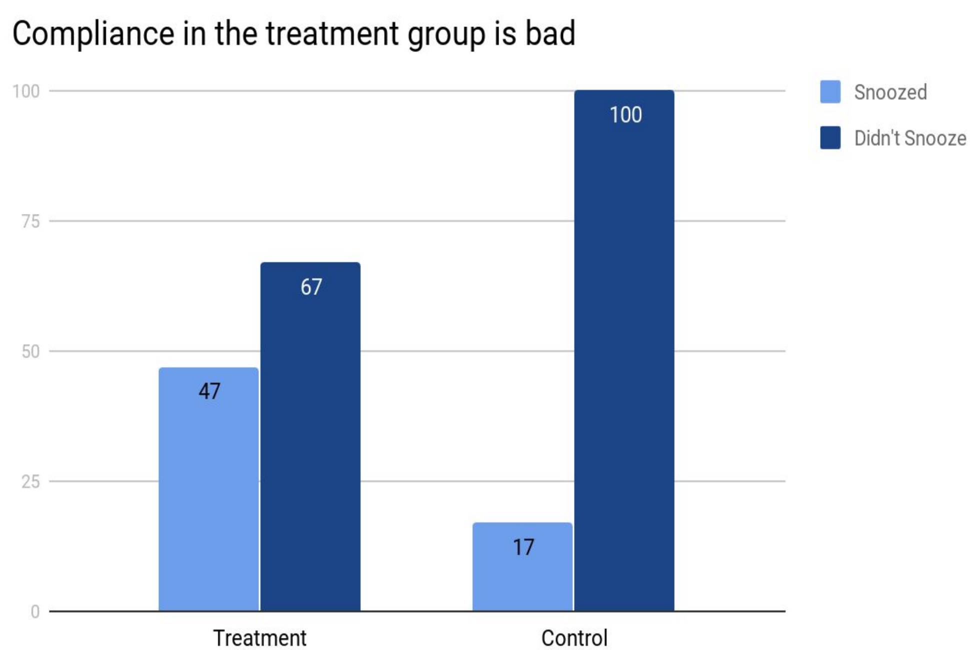

Figure 8: Compliance issues in the experiment

We see from the figure above that out of the 114 treatment observations, only 47 (41%) of the observations are compliers. In the control group, the state is much better. But on an overall level, it looks like most people (72% of overall sample) do not snooze alarms. This is quite different from the number of people who keep none or one alarm (55% according to pretreatment survey). Based on this, we will be calculating our model results on both - assigned groups of treatment and control, and the complier groups of treatment and control.

Regressions

For our experiment, we were able to collect longitudinal data for subjects who were randomly assigned to treatment or control groups during the 8 days that the experiment was conducted. In order to estimate the average treatment effect of snoozing alarms on the mental state of an individual (alertness and freshness) by accounting for within subject comparisons as well, we applied panel OLS regression to the following equation:

$$Y = \alpha + \beta(Treat) + \gamma\,(Covariates) + \delta + \mu$$

Here, Y is the outcome measure - which in our case would be measure of alertness or measure of freshness. α represents the intercept term, β is the causal estimate of average treatment effect, γ is set of additional coefficients for the covariates we use in our analysis, δ is known as the entity effect and is a feature of the Panel OLS regression and μ is the unobserved disturbance term.

We will be building two sets of models:

a) first with alertness as the outcome measure
b) second with freshness as the outcome measure

Looking at the distributions of both the outcome measures, we feel that there is a better chance of seeing a difference for the alertness model and hence that is the hypothesis we will test first. We will only then look at the models for freshness. We will use Bonferroni corrections for testing multiple hypothesis.

As discussed in the previous section on compliance, we are building/comparing two models for each outcome measure - one on the overall data (intent to treat) and second on the compliers (CACE)

Note: While using Panel OLS, we didn’t need to add the covariates which we had collected for each individual during the pretreatment survey since the effects of those covariates if captured in the entity effects term. We only add those covariates which change for each entity across time.

Model for Alertness

The following table shows the comparisons of these two models for the alertness model.

Model Comparison

	All data	Compliers
Dep. Variable	out_alert	out_alert
Estimator	PanelOLS	PanelOLS
No. Observations	256	163
Cov. Est.	Unadjusted	Unadjusted
R-squared	0.0622	0.1181
R-Squared (Within)	0.0622	0.1181
R-Squared (Between)	0.1882	0.1340
R-Squared (Overall)	0.0885	0.1239
F-statistic	2.7443	3.1069
P-value (F-stat)	0.0201	0.0114
Intercept	2.3195	2.4785
	(6.3999)	(5.0639)
C(activity_yesterday)[T.1]	-0.2817	-0.3442
	(-1.6470)	(-1.6220)
C(is_weekend)[T.1]	-0.1137	-0.1061
	(-0.8050)	(-0.5920)
treat	-0.0056	-0.4290
	(-0.0448)	(-2.4419)
hours_slept	0.1307	0.1032
	(2.7169)	(1.5815)
stress_yesterday	0.0886	0.1227
	(1.4736)	(1.5756)
Effects	Entity	Entity

T-stats reported in parentheses

Table 3: Panel regression model comparison for alertness model

The table above shows the T-stats in parentheses, and from the T-stats, we can see that see that the treatment effect is significant for the model built on compliers data. To get a clearer picture of the p-values and the standard errors of that model, we can check the summary of just the model made on complier data.

```
                        PanelOLS Estimation Summary
================================================================================
Dep. Variable:              out_alert   R-squared:                        0.1181
Estimator:                   PanelOLS   R-squared (Between):              0.1340
No. Observations:                 163   R-squared (Within):               0.1181
Date:                Tue, Dec 11 2018   R-squared (Overall):              0.1239
Time:                        18:33:58   Log-likelihood                   -188.75
Cov. Estimator:            Unadjusted
                                        F-statistic:                      3.1069
Entities:                          44   P-value                           0.0114
Avg Obs:                       3.7045   Distribution:                   F(5,116)
Min Obs:                       0.0000
Max Obs:                       7.0000   F-statistic (robust):             3.1069
                                        P-value                           0.0114
Time periods:                       8   Distribution:                   F(5,116)
Avg Obs:                       20.375
Min Obs:                       15.000
Max Obs:                       25.000

                                  Parameter Estimates
=============================================================================================
                           Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
---------------------------------------------------------------------------------------------
Intercept                     2.4785     0.4894     5.0639     0.0000      1.5091      3.4479
C(activity_yesterday)[T.1]   -0.3442     0.2122    -1.6220     0.1075     -0.7644      0.0761
C(is_weekend)[T.1]           -0.1061     0.1792    -0.5920     0.5550     -0.4610      0.2488
treat                        -0.4290     0.1757    -2.4419     0.0161     -0.7770     -0.0810
hours_slept                   0.1032     0.0652     1.5815     0.1165     -0.0260      0.2324
stress_yesterday              0.1227     0.0779     1.5756     0.1178     -0.0315      0.2770
=============================================================================================

F-test for Poolability: 1.6560
P-value: 0.0193
Distribution: F(41,116)

Included effects: Entity
```

Table 4: Panel OLS model summary for alertness model for compliers

From the table above, we can see that only the treatment variable has a statistically significant impact on alertness. But the coefficients of the other variables (although insignificant) also make sense. `stress_yesterday` is coded such that 5 means ‘Not stressed at all’ and 1 means ‘Very stressed’. We can see that `stress_yesterday` and `hours_slept` positively impact the alertness score, whereas `is_weekend` == 1 and `activity_yesterday` == 1 negatively impact the alert score.

Model for Freshness

The following table shows the comparisons of the two models (overall data and complier data) for the freshness model. Applying Bonferroni corrections, we would need to see a p-value less than 0.025 to get a statistically significant result.

Model Comparison

	All data	Compliers
Dep. Variable	out_refresh	out_refresh
Estimator	PanelOLS	PanelOLS
No. Observations	256	163
Cov. Est.	Unadjusted	Unadjusted
R-squared	0.1911	0.1595
R-Squared (Within)	0.1911	0.1595
R-Squared (Between)	0.0033	0.2112
R-Squared (Overall)	0.1332	0.1594
F-statistic	9.7821	4.4037
P-value (F-stat)	0.0000	0.0010
Intercept	1.0751	1.5787
	(3.5159)	(3.7403)
C(activity_yesterday)[T.1]	-0.1472	-0.2279
	(-1.0200)	(-1.2456)
C(is_weekend)[T.1]	-0.0800	0.0266
	(-0.6713)	(0.1719)
treat	0.0970	-0.1735
	(0.9195)	(-1.1452)
hours_slept	0.2007	0.1475
	(4.9443)	(2.6222)
stress_yesterday	0.1950	0.1672
	(3.8461)	(2.4893)
Effects	Entity	Entity

T-stats reported in parentheses

Table 5: Panel regression model comparison for freshness model

From the table above, we can see that the coefficients of treatment are not significant in either of the models as the T-stats reported in the parentheses is in (-1.96,1.96).

Power Calculations

Using Cohen's d, we calculated the effect size of the two sets of outcomes and showed the power with the current sample size and the expected sample size for 80% power

Outcome	N (Treatment Compliers)	Cohen's D Effect Size	Power	Expected Sample size for 80% power
Alert	47	-0.398	0.479	100
Refresh	47	-0.229	0.196	298

Table 6: Power Calculations

The calculations for these values are shown in the appendix.

It is clear from the table above that our experiment is highly underpowered. We need a lot more treatment compliers to achieve 80% power.

Limitations and Future Enhancements

Over the course of the experiment we have come to notice some limitations on what we can imply from the data we have gathered. We also have some thoughts on how we can overcome some of the limitations in follow-up analyses and experiments. We will first discuss the limitations of the experiment and then share our views on what could be some of the enhancements for future work.

Limitations

- The first limitation in the experiment design is that we will never be able to be sure how many subjects actually kept the alarms that were assigned to them since were are expecting them to administer the treatment
- Due to such low compliance numbers we have limited scope in calculating the Intent to Treat
- The treatment mails were presently sent 4 times a day to the 8 time zones. There was no automated method in place, since randomizations were happening at a daily cadence
- Currently, we had no way to ensure that subjects were responding to the mail as soon as the woke up. Late responses could highly bias the results we are seeing

Future Enhancements

- Our experiment instructions were only via emails. This may have played a part in attrition and non-compliance. Hence for future versions we could apply multi-channel approach of text, or best creating a custom alarm app where we could directly administer the treatment
- The problem of waking up before the alarm would still exist in such a scenario and therefore we might need to conduct such an experiment in a controlled environment where we can monitor the sleep activity of the participants
- We could look at incentivizing the participants in a way such that they are less prone to attrite during the longitudinal experiment.

Conclusion

To conclude, we can see that we need a lot more compliers to get a significant effect size for the treatment effect. But even from the limited complier data that we have, we can see a statistically significant negative effect of snoozing of size -0.43 (SE 0.17) on the alertness of an individual after he or she wakes up. The limited data shows no effect on the freshness of a person by snoozing alarms. This research direction is definitely interesting but the major non-compliance issues need to be solved before we can expect to see meaningful results.

References

1. https://www.weforum.org/agenda/2015/01/why-its-time-to-stop-hitting-the-snooze-button/
2. https://www.wsj.com/articles/sleep-experts-close-in-on-the-optimal-nights-sleep-1405984970
3. https://www.ncbi.nlm.nih.gov/pubmed/25130898
4. https://www.ncbi.nlm.nih.gov/pubmed/24260280

Appendix

List of covariates collected

Variable Name	Description
activity_yesterday	Whether subject did any rigorous activity previous night
age	Age of participant
avg_sleep	Average sleep time of subject
cat_alarm_nearby	Is the alarm near the subject at night?
cat_avg_alarms	Average number of alarms kept
cat_country	Country of subject
cat_deep_sleeper	Is the subject a deep sleeper?
cat_house_type	What kind of house does the subject live in?
cat_insomnia	Does the subject suffer from insomnia?
cat_kid_under_4	Does the subject suffer have a child under 4 years?
cat_noisy_roomies	Does the subject have noisy roommates?
cat_profession_group	What is the subject's profession?
cat_room_light	How much light enters the subject's room in the morning?
cat_traffic_noise	Does traffic noise come in the subject's bedroom?
hours_slept	How many hours did the subject sleep the previous night?
is_weekend	Is today a weekend?
stress_yesterday	How stressed was the subject yesterday?

Statistical Power Calculations

Power calculations

```
from statistics import mean, stdev
from math import sqrt
from statsmodels.stats.power import TTestIndPower
```

```
alpha = 0.05
power = 0.8
analysis = TTestIndPower()
```

Alertness Measure

```
from numpy import mean
from numpy import var

def cohend(d1, d2):
    # calculate the size of samples
    n1, n2 = len(d1), len(d2)
    # calculate the variance of the samples
    s1, s2 = var(d1, ddof=1), var(d2, ddof=1)
    # calculate the pooled standard deviation
    s = sqrt(((n1 - 1) * s1 + (n2 - 1) * s2) / (n1 + n2 - 2))
    # calculate the means of the samples
    u1, u2 = mean(d1), mean(d2)
    # calculate the effect size
    return (u1 - u2) / s
```

```
treat = df[(df.treat == 1) & (df.snooze ==1) ]['out_alert']
control = df[(df.treat == 0)& (df.snooze ==0)]['out_alert']

cohens_d = cohend(treat, control)
print("Cohen's D for Alertness measure = ", cohens_d)

p = analysis.power(cohens_d, nobs1= 47, ratio=1.0, alpha= alpha)
print("power = ", p)

result = analysis.solve_power(cohens_d, power= power, nobs1= None, ratio = 1.0, alpha = alpha)
print("sample size required = %.3f" % result)
```

```
Cohen's D for Alertness measure =  -0.39782442920634037
power =  0.479484397827908
sample size required = 100.156
```

Refresh measure

```
treat = df[(df.treat == 1) & (df.snooze ==1) ]['out_refresh']
control = df[(df.treat == 0)& (df.snooze ==0)]['out_refresh']

cohens_d = cohend(treat, control)
print("Cohen's D for Alertness measure = ", cohens_d)

p = analysis.power(cohens_d, nobs1= 47, ratio=1.0, alpha= alpha)
print("power = ", p)

result = analysis.solve_power(cohens_d, power= power, nobs1= None, ratio = 1.0, alpha = alpha)
print("sample size required = %.3f" % result)
```

```
Cohen's D for Alertness measure =  -0.22966012437024436
power =  0.1964794738029779
sample size required = 298.586
```