

Essay 2 Revised: Final Project Experiment Proposal

W241.2: Field Experiments

Ye (Kevin) Pang

[Google Doc](#)

Research Question

Does drinking a glass of warm milk before bed help you sleep better?

Background

A good night's sleep is vital to your health and ensures that you can start the next day refreshed and energized. According to a recent National Center for Biotechnology Information (NCBI) study, as many as 60 millions Americans have a sleeping disorder and people often will go to great lengths and try some oddball things to get a night of restful sleep.^[1] While there are many claims on different products to help people sleep easier/better, I want to use something simple, readily available, and easily accessible. Growing up, we've probably heard from our moms and/or grandmas that drinking a glass of warm milk before bed help us sleep better. Since it was easy to do, I am sure many of us simply complied. But, do you ever wonder how big is the effect (chemical + psychological) and is the effect sizable enough to make a causal claim? Using tools and techniques we learned from this class, I'd like to propose an experiment to study the causal claim and the average effect.

Subjects

The subjects for the experiment can be anyone who would agree if selected to drink a glass of warm milk 30 to 60 minutes before bed. For people who are lactose intolerant, we can still use lactose-free milk. However, our subjects will not include anyone with milk allergies due to medical issues. For the excludability assumption to hold, we also have to make sure that the subjects are not participating at other potentially conflicting/duplicating sleep related studies and hence are not taking medication or other treatments that may interfere with the treatment effect we want to test. Each subject will fill out a survey for us to collect basic information and helpful covariates. They will then be notified which group (treatment/control) they are in.

The subjects can be drawn from a variety of sources that include our class, volunteers from all sections of W241, MIDS program students, friends, families, and relatives. We can also use Amazon's MTurk if necessary to elicit more subjects. Ideally, we'd like to have more than 30 subjects. Each subject will also repeat the experiment for a period of 5-7 days.

Treatment & Variation

We will create variation through deliberate intervention of assigning subjects to either the control or the treatment. The treatment group consists of subjects that will drink a glass of warm milk before bed. The control group consists of subjects that will follow their own bedtime routines without consuming the glass of warm milk. Today, there are so many varieties of milk products to choose from, which makes the definition of “milk” a little bit ambiguous. Nonetheless, we think we can control for the variations in the choice of “milk” consumed by using covariates and/or group similar milk products into categories. For example, using group labels such as low fat, fat-free, coconut, soy, plant-based, etc. For people who drink customized milk^[3], we can account and accept them as long as it’s still passable as milk.

Outcome Measure(s)

The outcome measure/treatment effect we want to study is the qualitative word “better” in the research question. One way which “better” (or worse) can be quantified is using wearable/app assisted means^[2] to measure the time it takes to fall asleep. For people without access to those technologies, we can supplement with a survey where we have subjects answer a couple questions the following day. If both approaches are utilized, we will normalize the results so they are comparable. For example, we can use intervals instead of exact number of minutes.

Some of the secondary measures we can collect include other common sleep evaluation metrics such as duration of light and deep sleep cycles, and the time people went to bed and woke up. Along with covariates, the goal is to improve statistical power and precision with which the average treatment effect (ATE) is estimated.

Covariates

We will use covariates for two purposes: one is to reduce variance on the potential outcomes and the other is to assist in forming relatively homogenous groups for block randomized design. The covariates we will consider using should be related to the circumstances which the subjects are in before they went to bed: such as whether they exercised during the day, how tired were they, and what time they went to the bed. Other factors that may correlate with the treatment and impact affect how quickly one falls asleep include whether they sleep with a partner or if they live with young children. The whole list of potential covariates warrant a further discussion among the team members to make sure all necessary assumptions are reasonable. For each of the covariates, if it is categorical, we can also convert the levels into dummy variables.

Randomization

There are several approaches to randomization we can employ to remove selection bias. First, instead of simple (m/N) or complete (set m) randomization, we will opt for equal sized treatment and control groups. Both clustering and blocking can potentially work. Blocks are chosen to reduce variability. Clusters are usually subjects that have similar variability but may generate more sampling variability for the experiment. If there are natural clusters for our subjects, we will consider using clustering in addition to blocking. Otherwise, we will setup blocking on its own.

Statistical Analysis

We will use suitable statistical tests to study the average treatment effect difference (ATE) between the control and treatment groups. Some examples (but not limited to) include randomization inference, testing the sharp null hypothesis (p-value), calculating power, standard errors and confidence intervals.

Pilot

Conducting a smaller scale pilot will definitely help ensure the experiment design and data collection processes are robust and screen out any obvious issues. Before reaching out to other subjects, a pilot can mean running the experiment among the researchers ourselves. One of us can test the control group setup and another person can test the setup for the treatment group. Data collection pipelines as well as daily reminders should be setup to make sure our treatment subjects are compliant with what the experiment requires them to do.

Risks

1. It might be hard to do blocking beyond basic grouping factors such as gender and age. Other attributes of the individuals will be time consuming and hard to get.
2. The initial effect size has to be reasonably guessed
3. Sample size and time horizon of data collection will affect statistical significance
4. With survey based data collection, data quality will be susceptible to how accurately the subjects can recall the numbers. This in turn, will affect the experiment's statistical significance and study of the causal claim
5. App failures or inaccurate use of wearable devices can corrupt data, proper handling needs to be in place to get the subjects trained on how to use the technology and help them get back on track. Rigorous monitoring of incoming data quality will also help.

6. Wearables and apps may give slightly different numbers, need to understand those differences and also choose a more practical approach.

References:

1. <https://www.sleepadvisor.org/warm-milk-sleep/>
2. <https://www.azumio.com/s/sleeptime/index.htm>
3. <https://www.healthline.com/health/food-nutrition/warm-milk-to-sleep#1>