Nathaniel Velarde
W241.2 – Essay 2 Draft
September 26, 2018

Does Initial Starting Position on a Leaderboard Impact Subsequent Physical Activity Levels in Online Social Exercise Groups?

BACKGROUND:

Not only has the advent of internet-enabled fitness tracking apps and devices ("fitness trackers") has enabled individuals to track their physical activity in real-time, but it has also facilitated the sharing of this data with one's social network. Social comparison strategies which use rankings, leader boards, and/or social status markers to motivate increases in physical activity are a standard feature of fitness trackers. The user does not necessarily have to personally know the members of their group (e.g., Peloton real-time leaderboards), what is most important is knowing how they measure up to them.

The underlying concept is simple. Create a competitive environment where users work towards their goals individually, and differences in goal attainment will motivate individuals to adjust their aspirations, and ultimately, physical activity levels upward. Zheng et al., (2016) conducted a randomized controlled trial to identify what features of online social networks can increase physical activity and found that competition is effective for motivating increased physical activity.

RESEARCH QUESTION:

We propose to examine the question of whether initial starting position on an online social network physical activity leaderboard affects subsequent motivation levels as measured by physical activity levels. If you start at the bottom of your peer group, is there an "underdog" or "embarrassment" effect that motivates you to subsequently try harder relative to those who start at the top who could take their starting position as "permission" to coast? Or is the converse true, and those who find themselves at the top become more motivated to stay at the summit while those at the bottom succumb to defeatism? At this early stage, both hypotheses regarding the sign of the causal parameter seem equally plausible.

If initial starting position does have a casual (positive or negative) effect on subsequent physical activity then initial leaderboards could be "gamed" to positively jumpstart an individual or group fitness program. If so, a secondary question that could be pursued is how the long the effect persists.

EXPERIMENTAL DESIGN:

Subject Assignment to Treatment and Control Groups

To determine if initial starting position on a social group exercise group leaderboard can increase/decrease subsequent physical activity, we will conduct a 4-arm randomized control

trial. We plan on having at least 80 total subjects in the study (this is our working number at this stage which will be refined in later stages). The pool of potential subjects would be adults (> 18 years old) who have to access to a fitness device/app. The only exclusion criterion is that a subject is not sharing physical activity data with anyone (i.e., not a member of social exercise group). We plan on recruiting subjects via local college campus and/or craigslist.com. The subjects will receive nominal compensation for participating in the study, primarily for (a) their time in filling out demographic questionnaire and (b) granting the study access to their physical activity data, which will be sent to the study automatically via their fitness tracker, for the duration of the study, which will be approximately 2 weeks.

We will use block randomization, where subjects are partitioned by gender, to assign participants to one of four experimental conditions – "control", "bottom treatment", "middle treatment" and "top treatment" – which are described in more detail below. We are motivated to block by gender to reduce potential sampling variability – e.g., gender differences in daily physical activity levels and response to social comparison. We do not anticipate any natural clustering.

As of now, our experimental design calls for 20 subjects, comprised of 10 men and 10 women, to be placed into each experimental condition. However, given the number of treatments, we may need to increase the number of subjects per treatment to increase statistical power.

Treatments

The study participants will be told that the purpose of the study is examine the mix of physical activity performed by people on a daily basis – hence the need for fitness tracker data. All participants will be given a demographic questionnaire at the beginning of the study. The questionnaire serves two purposes: (a) collect data on potential covariates and (b) "sell" the guise to treatment subjects that their assigned peer group is as comparable to them as possible (remove circumstantial self-rationalizations for a subject's initial ranking, such as "I work longer hours relative to my peer group, therefore that explains my ranking."). The questionnaire will gather information such as age, location (city), relationship status, employment status, household demographics (number of underage and/or senior dependents) and work/school schedule (number of hours spent working/studying per week), outside commitments (volunteer hours per week), and current holder of a gym membership or regular fitness class attendee (Y/N).

Participants in the control group will not be placed in any online social network. Their physical activity data will be collected for two weeks and analyzed for comparison with the treatment groups. There will be no communication between the control subjects and the researchers during the first three days of the study. At the end of the third day, the subjects will be given a summary of their activity for the past three days. For the remainder of the study, the control subjects will be sent a summary of their daily activity. This pattern of communication will mirror that for the treatment groups in terms of timing and format. The differences in content are described in detail below for the treatment groups.

For the first three days of the study, the treatment subjects' physical activity data will be collected to establish a baseline level of physical activity. At the end of the third day, the treatment subjects will be informed that based on the demographic questionnaire, they have

been placed in a 5-member online social fitness group and given the initial leaderboard for their group. This leaderboard will display each group member's name/ID, group rank and average physical activity statistics based on the past three days. Unbeknownst to the treatment subjects, their initial ranking and their fellow group members (and their respective physical activity numbers) will be fictitious. The treatment subjects' initial ranking will be determined by their assigned treatment group.

Subjects in the "bottom" treatment condition will find themselves at the bottom (5th) of the initial leaderboard. The physical activity metrics of their peers will be set such that the gap between 5th and 4th position is 10% and each subsequent position above 4th is randomly determined to be between 1% and 5%. At the end of study days 4-14, the bottom treatment group subjects will be sent an updated leaderboard. While the subject's activity data will be "real", the day-to-day performance of the subjects' peer group members will be simulated using physical activity data collected on all study subjects for the first 3 days. The simulated peer group members' activity will be made to resemble a random walk (prior day's activity level + noise factor based on the variance of actual study group daily activity).

The treatments for subjects in the "middle" and "top" treatment conditions are similar. They differ only in initial starting position. The "middle" treatment group subjects will be given an initial rank of 3. The physical activity metrics of their peers will be set such that the gap between 3rd and 2nd will be set at 10% and that 4th will also be 10% below 3rd. The gaps between 1st and 2nd and 4th and 5th will be randomly determined to be between 1% and 5%. The "top" treatment group subjects will be given an initial rank of 1. The physical activity metrics of their peers will be set such that the 1st will 10% above 2nd. Each subsequent peer position below 2nd is randomly determined to be between 1% and 5%.

MEASURES

The outcome of interest will be the average change in daily physical activity relative to the average three-day baseline measurement period post the revelation of the subjects' initial peer group rank. Fitness trackers collect different types of activity – steps, floors climbed, distance run, etc. One of our challenges will be develop a standardized, apples-to-apples, metric for physical activity to serve as our ranking and outcome measure.

STATISTICAL ANALYSIS

We will use our experimental data to calculate estimates of the Average Treatment Effect ("ATE") and standard error of the ATE, for each treatment condition and their respective gender blocks. We will test the sharp null hypothesis of no treatment effect via randomization inference and t-tests.

PILOT STUDY

We plan to run an initial pilot study involving up to 10-20 subjects. The goal of the pilot study is to work through the mechanics of data collection – setting up automated data transfer from the subjects' fitness trackers – and leaderboard design. This pilot study will enable the researchers to prevent potential issues from asymmetries in measurement arising from differences in how information is presented to subjects in different experimental conditions as well as the relative

timing of communications to different groups. Ideally, all groups should receive as similar information (treatments) at approximately the same time.

WHAT MIGHT GO WRONG?

This experiment's goal is determining whether the initial starting position on a social fitness group leaderboard has an impact on subsequent physical activity levels. Under the experimental design presented here, there is the possibility of confusing the impact from the subjects learning their initial position (and gap between themselves and their closest peer – set at 10%) with the informational impact of their changing relative positioning in days 4-14. The experimental design tries to control for this by making the simulated peer group performance random for each subject for days 4-14. The thinking being any (positive/negative) informational cues from how peers do subsequent to the initial leaderboard will be averaged out over the treatment subjects as the peer data is randomly generated. One possible (and simplifying workaround) is to provide subjects with only the initial leaderboard on Day 3 and the final leaderboard on Day 14 (which would be meaningless, but useful to mentally close the loop with the study participants), to remove any confounding effects caused by daily leaderboard updates, which could have a confounding motivating/demotivating effect depending on the subject's relative performance for that day.