

Survey into the Accuracy of Perception of Minutes Active Per Day

Team Survey & Design Project - Consumer Electronics & Internet Group

Tamara Williams, Zachary Urgena, Spencer Li

Northwestern University

Team Survey & Design Project - Consumer Electronics & Internet Group

Survey into the Accuracy of Perception of Minutes Active Per Day

Recent studies reveal that people typically underestimate the amount of activity they do in a day (Dyrstad, Hansen, Holme, & Anderssen, 2014). As part of a continuing program to improve the health of the university population by promoting exercise, we are conducting a study with the university community. The goal is to understand any discrepancy between perception and actual levels of activity. These results will inform future programs to motivate and incentivize the community to become more active as part of their daily routine.

Methodology and Mechanics of the Study

Approach

Given the ubiquity of cell phones among the faculty, staff and students, the we will rely heavily on the use of cell phones. Through a custom phone application (“app”), actual activity will be tracked, then compared to the self-reported values from study participants. For privacy, personally identifiable information (PII) will be encrypted. Tracking will be done by a globally unique identifier (GUID) assigned to each study candidate. Outbound communication will be done as push-notifications in the phone app. Notifications will be sent via an automated process, creating a firewall between the participants and the researchers.

Participants

Sample frame and methodology.

The frame population for this study is the students, staff and faculty of the university. There are known gender-based differences in patterns of activity (University of Exeter, 2009) and known age-related differences in activity (Bélanger, Townsend, & Foster, 2011). To account for these known differences, we are going to use stratified sampling (Groves, et al., 2009). The strata to be used are shown in Table 1. Primary strata will have proportionate allocation of the 20,000 people in the data frame.

We are using a 95% confidence interval with a 5% margin of error along with a population parameter of 20% of adults who exercise (Jaslow, 2013) to determine that the sample size needed for our population is 240 participants. The equation used to determine the sample size is:

$$n = \frac{Z_{CI}^2 p(1-p)}{E^2}$$

To have a suitable sample size, we are setting a minimum 2% or better participation rate across the university, which totals 400 (larger than the minimum calculated above). Given the strata breakdown in Table 1, we will need 100 male and 100 female participants between the ages 18-29, 60 male and 60 female participants between the ages 30-49, and 40 male and female participants over the age of 50. Studies show an average response rates to employer surveys in the 57% range (Barush & Holtom, 2008) and students respond to web-surveys at an average 31.5% rate (Sax, Gilmartin, & Hagedorn, 2008). Since we are drawing on students and employees of the university, it is expected that we will have no trouble meeting the minimum target of at least 2% of the population.

To maintain proportion, the strata with the lowest participation rate will gate the number of participants across all strata. Should any strata have fewer than 5 participants, the study will not go forward. The secondary strata will be collected and processed to see if there are significant aggregate differences. Secondary strata will not be proportionate and will not gate the study. Historically, the university has not collected “decline to identify/other” gender data so proportions are not known.

Recruiting Participants.

To recruit people into the study, we will leverage the established campus communication channels. We will run ads in the campus newspaper, post ads on the various bulletin boards, reach out directly to student and faculty groups, use text solicitation as permitted under university policy, and make social media posts in relevant venues seeking participants.

Interested parties will go to a website and fill out the in-take survey form (Appendix A). Form data will be used to assign participants to the correct stratum, and establish their eligibility for the study. The backend processing of the form will also assign a GUID as a key-value part with the email of the participant. Email addresses will be encrypted and all tracking will be done via the GUID.

Participant Selection.

Since portions of the study will be conducted via a smartphone app, respondents to the initial in-take survey must own a phone that can run the app, and be willing to have the phone with them during a significant portion of the day. As part of the app installation process, privacy consent will be sought. Anyone declining consent will not be able to install the app.

To be eligible to participate, respondents must meet several criteria. They must carry their phone at least twelve hours per day. The in-take survey asks about current carrying behavior to give preference to respondents who already carry their phone for significant periods of time. We are asking them to estimate total daily activity, and recording less than twelve hours of actual data will not allow for an accurate comparison of estimated versus real activity levels. They must have a phone capable of running the study app which records actual activity as a function of the phone's accelerometer data. To be eligible for participation, respondents must also indicate they are willing to run the app on their phones, and willing to have their data collected and analyzed. If more people register than are needed for a stratum, participants will be selected at random from the pool of eligible respondents. The goal is to have the largest possible number of people participate within the framework of the strata.

Study Mechanics

There are four phases to the study. First is the recruitment and selection process detailed above. Second, an initial survey (Appendix B) establishes a baseline for each selected participant. Third, the data collection phase, participants will carry their phones with the app installed, every day for the duration of the two-week study period. Evaluation of the data will be

the final step. During phase four, participants will be able to access the actual values for their reported active minutes per day.

Contingency Planning

With relying on volunteers participating, there are some confounding factors that may not be directly controlled, affecting our data collection. Initially, one of the main requirements to proceed is having an appropriate response rate. If we fall under our necessary number of participants, this would greatly affect our collected data leading to possible non-response bias. If this situation occurs, depending on resources, we would either need redesign our initial survey or develop incentives. As estimated from our target sample and desired response rate, previous studies indicate we have given enough margin to safely meet this minimum.

Another possible issue is insufficient usage of participants' mobile devices during the study or quitting in the middle of the study. This can happen due to forgetting their phone, losing power, or deciding to stop participating. In the event, there is missing data or quitting, we would need to adjust for the final analysis. However, for any outlier data produced from these occurrences, it would still be included in the analysis. To mitigate these issues, use push notifications to ensure the study is kept in mind for the participants. This will continually inform them to carry it during their activities but also making sure they sufficiently charge the device. Also, with high "perceived length of time or complexity of self-administered instruments reduces cooperation" (Groves, et al., 2009) so to reduce quitting during the study, removing any manual interactions will improve retention.

Data Processing, Cleaning and Outliers

It is possible that some data may have issues. Phones run out of power, or are left at home, which will result in missing, or incorrect data. Where data are totally missing, we will drop that record from the individual's personal collection, and adjust the calculations to accommodate a reduced number. In the case of inaccurate data, we have no insight to tell us it is

inaccurate. If it sufficiently “off,” as in the case of a forgotten phone, it will present as an outlier in the data set.

Data collection is tightly constrained via the use of web forms. Data are checked at the time of form submission for correct formatting. A participant will not be able to submit a form with incorrectly formatted data. Additional cleaning will not be required.

A graphical representation of the data will be generated, in addition to quantitative values. Outliers will be flagged on the visualization, but will not be dropped from the processing. It is not possible to know if the outlier is caused by a collection error (forgotten phone, dead battery) or is a true reading, so the data will be allowed to stand.

Potential Bias

Despite best efforts to design a survey that is clearly understood and fair, biases are difficult to completely avoid. An area in this study which may present bias in the participants is in the in-take survey, where gender and age identification are requested. Known activity differences exist between males and females (University of Exeter, 2009) and those in different age groups (Dyrstad, Hansen, Holme, & Anderssen, 2014). That knowledge may transfer into the fulfillment of the perceived stereotype (Schmader, 2002), i.e. a female may provide lower estimates of her daily activity, or someone who is fresher into adulthood may overestimate their daily activity. These questions cannot be omitted, since they are key strata for this study. Another potential for bias comes from the framing of our survey, which seeks to capture information for a health-related initiative. This may influence our sample to have a disproportionately high number of health-focused individuals. We will continue to with the study with the knowledge of the potential biases that may arise from the questions addressed above.

Data Analysis

For each of the strata we will calculate a per-person accuracy scores as follows:

$$daily\ accuracy_p = \frac{\sum_1^{14}(Estimate_i - Actual_i)}{Actual_i} \text{ and } mean\ accuracy_p = \frac{\sum daily\ accuracy_p}{14}$$

The mean for each stratum will then be computed as: $\bar{x}_h = \frac{\sum_1^n mean\ accuracy_p}{N_h}$

To arrive at the mean for the accuracy of people's self-assessment of daily activity across the target frame, the strata must be combined (Groves, et al., 2009). The frame mean will be

$$\text{calculated as: } \bar{x}_{st} = \frac{\sum N_h \cdot \bar{x}_h}{N}$$

where N_h is the number of people in each stratum.

From the strata mean accuracy, the combined frame mean, and the variance, we will be able to answer the question, "How accurate are members of the university at estimating their activity level per day?" We will be able to answer the question for each stratum, and for the whole university population. Data from the secondary strata will also be processed using the same methods. Results from the secondary strata will be compared to the entire population, if there are any significant differences, they may result in additional separate studies.

Conclusion

In our world where sedentary lifestyles are becoming more commonplace, it is important to work towards a healthy lifestyle. We hope to find valuable trends in our data that shine light on how accurately people measure their levels of activity. The knowledge gained from this study may influence individuals to be more aware of their own biases and invest in equipment that can objectively measure their daily activity.

References

- Barush, Y., & Holtom, B. C. (2008, August). Survey response rate levels and trends in organizational research. *Human Relations*, 61(8), 1139-1160.
- Bélanger, M., Townsend, N., & Foster, C. (2011, March). Age-related differences in physical activity profiles of English adults. *Preventive Medicine*, 52(3-4), 247-249.
- Dyrstad, S. M., Hansen, B. H., Holme, I. M., & Anderssen, S. A. (2014, January). Comparison of Self-reported versus Accelerometer-Measured Physical Activity. *Medicine & Science in Sports & Exercise*, 46(1), 99-106.
- Groves, R. M., Fowler Jr., F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology, Second Edition*. Hoboken, NJ, U.S.A: John Wiley & Sons.
- Jaslow, R. (2013, May 03). *CDC: 80 percent of American adults don't get recommended exercise*. Retrieved March 12, 2017, from <http://www.cbsnews.com/news/cdc-80-percent-of-american-adults-dont-get-recommended-exercise/>
- Sax, L. J., Gilmartin, S. K., & Hagedorn, L. S. (2008). Using web surveys to reach community college students: An analysis of response rates and response bias. *Community College Journal of Research and Practice*, 32(9), 712-729.
- Schmader, T. (2002). Gender Identification Moderates Stereotype Threat Effects on Women's Math Performance. *Journal of Experimental Social Psychology*, 38(2), 194-201.
- University of Exeter. (2009, January 5). *Studies reveal lifelong gender difference in physical activity*. Retrieved from EurekaAlert!: https://www.eurekaalert.org/pub_releases/2009-01/uoe-srl010509.php

Tables

Table 1

Study Strata as a proportion of the population, number of people in the stratum, and desired number of participants

Age	Male	Female	Decline to identify/Other
18-29	.25 [5000] {100}	.25 [5000] {100}	
30-49	.15 [3000] {60}	.15 [3000] {60}	
50+	.10 [2000] {40}	.10 [2000] {40}	
Decline to identify/Other			

Note: The 6 shaded strata are the primary focus for the study, the rest are secondary. The fractions shown in Table 1 are the proportion of the University community in that stratum, based on demographics from Human Resources and the Registrar's office. The brackets [] indicate the corresponding number of potential participants. The braces {} are the desired minimum number participants, which equates to a 2% response rate.

Appendix A

In-take survey to establish eligibility for study participation and identify stratum membership

Membership

Thank you for participating in this study. We are conducting research as a health initiative for the university. Please fill out these questions to the best of your ability.

Section I

Do you own any of the following? (select all that apply)

- ☐ iPhone 4 or later
- ☐ Android phone (Samsung Galaxy S, or later)
- ☐ Google Pixel phone
- ☐ I do not own a smartphone

How many hours a day (on average) do you carry your phone?

- ☐ 0 - 4
- ☐ 4 - 8
- ☐ 8 - 12
- ☐ 13 or more

Are you willing and able to carry your phone with you a minimum of 12 hours a day, every day, for the entire length (2 weeks) of the study?

- ☐ Yes
- ☐ No
- ☐ Maybe or can't decide

Are you willing and able to install a custom app and allow its data to be sent to the research team for the entire length of the study?

- ☐ Yes
- ☐ No
- ☐ Maybe or can't decide

Are you willing to have your data collected and analyzed for this research?

- ☐ Yes
- ☐ No
- ☐ Maybe or can't decide

Section II

Do you gender identify as:

- ☐ Female
- ☐ Male
- ☐ Prefer not to say/Other

Your age range:

- ☐ 18 - 29
- ☐ 30 - 49
- ☐ 50 or older
- ☐ Prefer not to say/Other

Your participation is very valuable. Thank you for taking our survey.

Appendix B

Survey to collect baseline data from study participants

Form is a web-based collection instrument, and can be filled in via PC or phone.

Please fill in the table below with total of your estimates for your daily number of active minutes:

Day of the week	Total Active Minutes (whole number)	Percent confidence (percentage)
Monday		
Tuesday		
Wednesday		
Thursday		
Friday		
Saturday		
Sunday		

Your participation is very valuable. Thank you for taking our survey.

Appendix C

Daily Survey

Survey data collection instrument to capture participants' daily estimates of their activity. Form is filled out by the participant on their phone.

How many total minutes do you estimate you were active today? Enter a whole number.

How confident are you in that estimate? Enter as a number from 0 – 100

Your participation is very valuable. Thank you for taking our survey.