

# Effects of Health-Focused Messages on Physical Activity

Alexander Bendeck

## Introduction

### Study Background

The Duke Motivated Cognition and Aging Brain Lab (MCAB Lab), affiliated with the Psychology and Neuroscience departments, conducts research into how to motivate adults to be more physically active. To this end, the lab is conducting a study where participants are given a FitBit to track their number of daily steps. After a brief period where baseline data is collected, participants begin receiving a text message (referred to as a “health message”) every day which discusses the importance of physical activity. The order of the messages was randomized at the beginning of the study and is consistent for all participants. Participants receive a message every day for 80 days, and also fill out a survey each day to indicate (i) how relevant they find that day’s health message and (ii) their general feelings of “self-efficacy” (confidence in their ability to control their own behavior and environment) that day.

Each health message is labeled as either a “social” message which discusses the implications of the participant’s physical activity and healthy living on loved ones, or a “nonsocial” message which eschews these themes. Similarly, each message is also labeled as “positive” or “negative” depending on whether it frames its content in terms of the benefits of physical activity or dangers of inactivity, respectively. For example, one “negative, social” message used is: “Physical inactivity can make daily tasks harder to perform, leaving you with no energy to spend time with friends and family.”

### Goal

My overall goal is to use the data that has been collected thus far in order to draw conclusions about the relevance of health messages to adults as well as the impact of health messaging and self-efficacy on physical activity. I am primarily interested in the effect of the type of message received (i.e. “positive social”) on self-reported message relevance and daily physical activity, as well as whether there is a discrepancy between the two.

### Significance

The conclusions from this analysis will have implications both for future research and policy. If this study finds that certain types of health messages are effective at motivating adults to be more physically active, future studies could build on this work by designing even more personalized messages for individual adults. Such findings could also inform future decisions by policymakers and nonprofit organizations as they try to promote healthier living as a method of preventive health care. Even on a more personal level, the findings of this study could simply help people understand how to talk to their loved ones about their health in a way that will motivate them to make positive lifestyle choices.

## Scope and Related Work

The questions I will be addressing in this analysis do not include all of the research questions for this study. The study also has a neuroimaging component, where participants read health messages in an fMRI scanner so that their brain activity can be analyzed by researchers. Other undergraduate students' goals on this study rely on this neuroimaging data to understand participant behavior based on their brain scans. I will not be addressing any questions regarding fMRI data, nor is the brain imaging data present in the dataset I will be using. In this way, my analysis and research goals are disjoint from other ongoing work for this study.

In the literature, other studies have served as a proof of concept for the basic study design of using text messages to encourage healthy lifestyle choices; in a 2009 study, 79% of participants said that being sent an average of three health-focused text messages per day helped them towards their weight loss goals (Gerber et al. 2009). Recent work in the Psychology/Neuroscience community has also shown that adults are often more motivated by social rewards and positively-framed messages as opposed to socially irrelevant or negatively-framed messages (Mikels et al. 2016). Studies have also commonly used a record of steps taken as a proxy for overall physical activity (Kraus et al. 2019, Pillay et al. 2015). Such studies have used modeling techniques such as Analysis of Covariance (Pillay et al. 2015) and multiple-sample structural equation models (Chaumeton et al. 2010). Analyzing survey-based data in related studies has been done with little statistical sophistication, only computing simple percentages of participants who agree or disagree with certain statements (Gerber et al. 2009).

## Data

### Description

Each row in the dataset corresponds to one calendar day for one individual study participant. There are 36 unique participants in the study, each of whom has over 90 rows for different calendar days. Participants received 80 days of health messages; data was collected for a few days both before and after the messaging period.

As mentioned previously, I am interested in both the participants' self-reported ratings of message relevance as well as their daily physical activity (number of steps taken) tracked by a FitBit. I will investigate using both of the following as *response variables* in different models: (1) the participant's self-reported *rating of the personal relevance of the health message* received that day from 1 to 10 (NA if no message received that day), and (2) the participant's *total steps taken* over the course of that day.

*Predictor variables* of interest include the following: (1) the *age* of the participant in years; (2) the *gender identity* of the participant, chosen from "Male", "Female", or "Other" (though "Male" and "Female" were the only genders reported by participants); (3) a binary indicator of *whether a health message was received* that day; (4) a categorical variable reporting whether the health message received that day was a "*positive*" or "*negative*" message (NA if no message received that day); (5) a categorical variable reporting whether the health message received that day was a "*social*" or "*nonsocial*" message (NA if no message received that day); and (6) the participant's self-reported *feelings of self-efficacy* that day on a scale from 1 (low) to 10 (high). Additionally, a *Subject ID* variable will be used for a random effect in hierarchical models.

## Exploratory Data Analysis

Overall, the dataset contains 4897 total data rows across 36 different participants. The table below shows a basic breakdown of the participants by demographics and the response variables of interest. The gender groups are reasonably balanced and have a similar mean age. A simple t-test (included in the Appendix) shows no significant difference between ages across gender groups, so doing comparisons between the groups is reasonable. On a daily basis, members of both gender groups also have similar numbers of steps and ratings of message relevance, though females tend to give slightly higher ratings. Histograms of the response variables, which are omitted here for conciseness (see Appendix), show data that is normally distributed with just a slight skew.

Table 1: Steps and Relevance Ratings by Gender

gender	subjects	mean age	mean steps	mean relevance rating
Female	20	49.65	9844.18	6.69
Male	16	44.25	10158.79	6.25

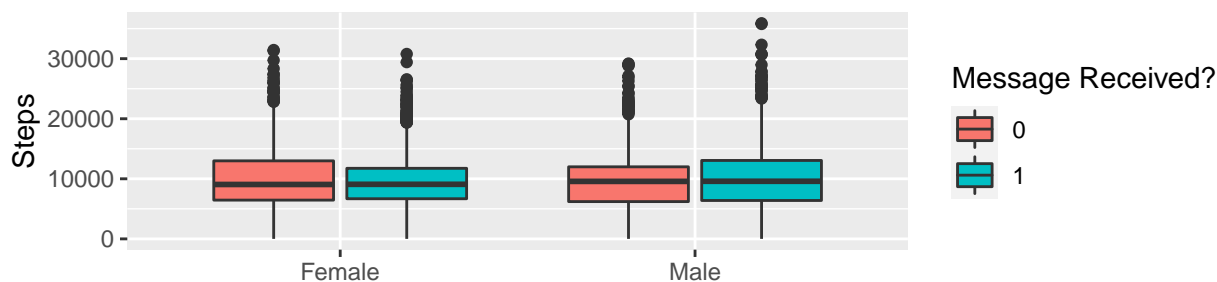
When comparing participants over or under the age of 50, there is not a huge difference in message relevance ratings, but perhaps average steps per day are a bit higher because more of these participants are retired and spend more time engaging in activities they enjoy rather than being at work.

Table 2: Steps and Relevance Ratings by Age Group

age group	subjects	mean steps	mean relevance rating
50 and younger	21	9664.14	6.44
51 and older	15	10330.24	6.50

Figure 1 below shows that total steps taken does not seem to vary based on whether or not a message was received. This is true across genders. This may indicate that there is no significant difference between physical activity on days when health messages are received compared to when messages are not received.

Figure 1: Total Steps vs. Gender, by Message Status



Similarly, Figure 2 additionally shows that on days when a message was received, total steps taken does not seem to vary based on whether the message was positive or negative. This is true across genders. This may indicate that there is no significant difference between physical activity on days when positive health messages are received compared to when negative messages are received.

Figure 2: Total Steps vs. Gender, by Valence

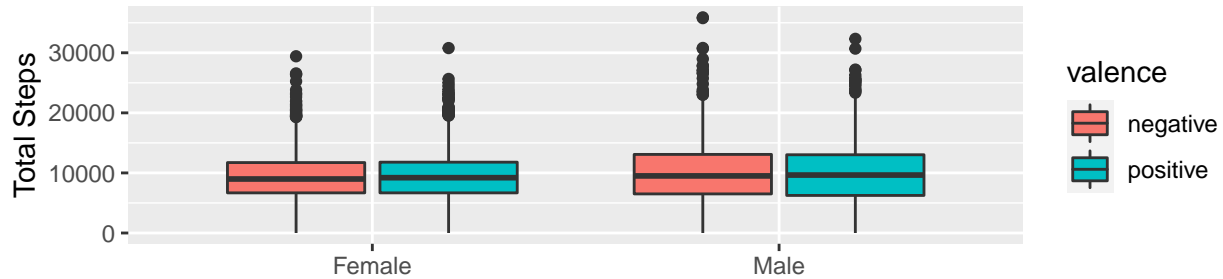
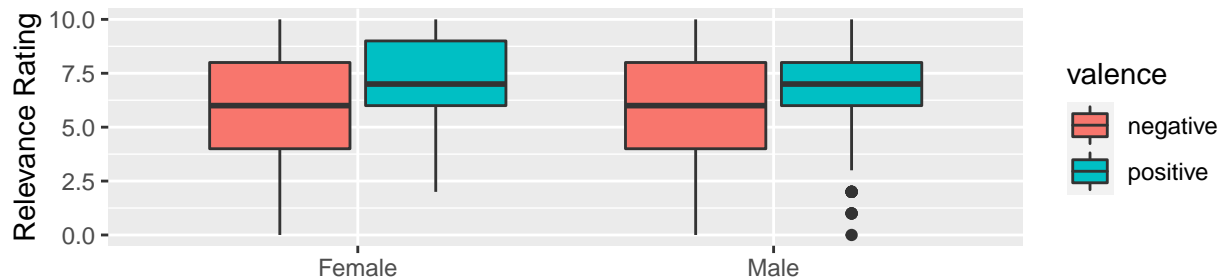


Figure 3 below shows that ratings of message relevance seem to vary based on whether a message is positive or negative. This effect appears slightly larger among females. This may indicate that adults find positive messages more personally relevant than negative messages, in accordance with the findings of Mikels et al., even if this might not manifest itself in increased physical activity. Note that we do not consider whether or not a message was received here, because they must have received a message that day in order to have provided a rating of relevance for the message.

Figure 3: Relevance Rating vs. Gender, by Message Valence



In order to see if the results here were being skewed by differences in individual tendencies with respect to steps taken or ratings given, I created “normalized” versions of the total steps and message relevance rating variables by subtracting each participant’s steps and ratings from their averages for that participant and then dividing the standard deviation. (For example, all data rows for subject 1011 would have total steps subtracted by the mean steps taken by subject 1011 across all days, and then this would be divided the standard deviation of total steps taken by subject 1011 across all days.) Plots produced with these normalized variables look generally the same and are provided in the Appendix for completeness. Normalized variables were not used in the models described below.

As a final note, from both a line plot of subjects and an autocorrelation function plot, I also discovered potential correlation between subsequent days for measurements of total steps as part of my Exploratory Data Analysis (see Appendix for plots). I will investigate this autocorrelation in my models.

## Modeling Number of Steps Taken

### Methods

Since different subjects may have very different numbers of steps taken per day on average and data observations for a given subject are likely to be correlated, I fit several mixed-effects linear regression models with total steps as a response variable and subject as a random intercept (see Sensitivity Analysis below for other models). Here I describe my final model with gender, age, self-efficacy, and type of message received (if any) as fixed effects. In order to address the potential autocorrelation of steps taken between subsequent days, this model also includes two autoregressive predictors: the number of steps taken in the two days preceding any given data row. For days 1 and 2 of the study, I used baseline data from before the study began for the autoregressive data. The model equation is below.

$$y_{ij} = \beta_0 + \alpha_j + \beta_1 \text{genderMale}_j + \beta_2 \text{age}_j + \beta_3 \text{self-efficacy}_{ij} + \beta_4 \text{msgNegNonsoc}_{ij} + \beta_5 \text{msgNegSoc}_{ij} \\ + \beta_6 \text{msgPosNonsoc}_{ij} + \beta_7 \text{msgPosSoc}_{ij} + \beta_8 y_{i-1,j} + \beta_9 y_{i-2,j} + \epsilon_{ij}$$

where  $\alpha_j \sim N(0, \sigma_\alpha^2)$ ,  $\epsilon_{ij} \sim N(0, \sigma^2)$ ,  $j$  is the subject index from 1 to 36, and  $i$  is the observation index for a specific subject (starting at 1 and usually ending between 90 and 100, depending on the number of observations for the subject).

### Results

Table 3: Coefficients for Steps Model

	Estimate	Std. Error	95% CI Lower	95% CI Upper
(Intercept)	7777.8643	776.4804	6224.9035	9330.8251
male	37.2155	1123.9157	-2210.6158	2285.0469
age	-0.6181	32.9682	-66.5544	65.3183
self efficacy	431.7071	32.0052	367.6967	495.7174
negative nonsocial	-301.0811	207.2073	-715.4957	113.3335
negative social	-513.0289	207.4302	-927.8893	-98.1684
positive nonsocial	-505.2327	207.5352	-920.3031	-90.1623
positive social	-264.5052	208.2347	-680.9746	151.9641
1-day lagged steps	0.1555	0.0174	0.1208	0.1903
2-day lagged steps	0.0860	0.0173	0.0514	0.1206

The fitted residual variance  $\sigma^2$  is 13213104, the random effect variance  $\sigma_\alpha^2$  is 10786383, and the intraclass correlation  $\frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma^2}$  is 0.45. (These variances seem large, but their square roots are in the range of 3500, which is not an unreasonable standard deviation when participants generally take around 10000 steps per day based on the Exploratory Data Analysis.)

Surprisingly, according to this model, it looks like receiving some types of messages actually made people take fewer steps than they would have compared to the baseline of not receiving a message at all. All else being held constant, receiving a negative social message or a positive nonsocial message reduces a participant's expected number of steps by around 500 (with a 95% confidence

interval from around -925 steps to -100 steps in both cases). For other message types, the negative effect of receiving the message on number of steps taken does not appear to be significant. Age and gender are not significant predictors of number of steps taken.

Self-efficacy appears to impact steps taken in a significant way; if someone rates themselves one point higher on the self-efficacy scale, we expect them to take around 430 more steps (with a 95% confidence interval of 368 to 496 added steps). The autoregressive coefficients are significant, although they do not have a huge effect. If yesterday a subject took 1000 more steps than their average, their expected number of steps today would increase by around 155. If two days ago they took 1000 more steps than their average, their expected number of steps today would increase by around 86.

This model shows no concerning pattern in its residual plot, which is included in the Appendix.

## Modeling Message Relevance Ratings

### Methods

It appears from the model above that receipt of a message, regardless of its type, did not cause participants in this study to take more steps. In spite of this, I am also interested in whether or not participants rate different messages as being more relevant than others.

My first approach to answering this question was to fit a linear mixed-effects model with message relevance rating as the response variable. Unlike the models for steps, these models only utilize data rows where a message was received (because otherwise there is no message rating, since there was no message to rate that day). The model (output in Appendix) found that holding all else constant, positive messages were correlated with higher message ratings than negative messages. While social messages showed no significant difference from nonsocial messages when the message is negative, a social component seems to be correlated with lower message ratings for positive messages. Age and gender had no effect. While this model provided some basic results to think about, it is clear that since the response variable is an ordinal rating between 0 and 10 (inclusive) it is not a natural fit for linear regression.

In an effort to find a model that fit the data more naturally, I created a new binary response variable to indicate whether or not a rating is higher than the average of all ratings over all days, which is around 6.5. Since ratings are whole numbers, this means that all ratings of 6 and lower get a 0 indicator variable and the ratings 7 and higher get a 1. I then fit a hierarchical logistic regression model with this indicator as the response.

$$\text{logit}(\mathbb{E}[y_{ij} | b_j]) = \beta_0 + b_j + \beta_1 \text{male}_j + \beta_2 \text{age}_j + \beta_3 \text{positive}_{ij} + \beta_4 \text{social}_{ij} + \beta_5 \text{positive:social}_{ij}$$

where  $b_j \sim N(0, \sigma^2)$ ,  $j$  is the subject index from 1 to 36, and  $i$  is the observation index for a specific subject (starting at 1 and usually ending between 90 and 100, depending on the number of observations for the subject). Since a message is always received in this model, it made more sense to have separate predictors for whether a message is “positive” or “social” plus an interaction term rather than using one combined message type predictor like in the steps model.

## Results

Table 4: Coefficients for Message Ratings Model

	Estimate	Std. Error	95% CI Lower	95% CI Upper
(Intercept)	-0.0891	0.2546	-0.5983	0.4202
male	-0.2708	0.3597	-0.9902	0.4486
age	0.0112	0.0107	-0.0103	0.0327
positive	1.0676	0.1305	0.8067	1.3286
social	0.0768	0.1274	-0.1780	0.3316
positive:social	-0.3435	0.1821	-0.7077	0.0207

The fitted random effect variance  $\sigma^2$  is 0.94, and the intraclass correlation  $\frac{\sigma^2}{\sigma^2 + \frac{\pi^2}{3}}$  is 0.22.

Like the linear regression model, this model also found that holding all else constant, positive messages were correlated with higher message ratings than negative messages. For instance, for a given subject and holding all else constant, receiving a positive message compared to the baseline of a negative message multiplies the odds of giving an above-average rating by  $e^{1.07} \approx 2.9$  (with a 95% confidence interval of  $e^{0.81} \approx 2.25$  to  $e^{1.32} \approx 3.74$  multiplicative change in odds). Social messages technically showed no significant difference from the baseline of nonsocial messages regardless of whether the message was negative or positive. However, the effect of a social component on positive messages is very nearly significant, so it is possible that a social component reduces the likelihood of an above-average rating for positive messages only. Age and gender had no significant effect on message relevance ratings.

The binned Pearson residuals (included in the Appendix) generally look good, so this model seems to be a decent fit for the data.

## Model Validation

To validate my model for number of steps taken, I removed Day 78 of the study from every participant, then trained the model on the dataset with Day 78 removed and used the remaining data to make out-of-sample predictions. (I originally wanted to use the last day of the study, Day 80, but several participants were missing data for that day.) For these predictions, the root mean squared error was 2963.75. Most days, participants took somewhere around 10000 steps (from Exploratory Data Analysis), so this error is not too large compared to the number of steps taken. Additionally, a plot of the residuals of these predictions against their predicted values (included in the Appendix) shows little of concern besides a couple of points with high predicted values and extraneously high residuals.

For message relevance ratings, I once again removed Day 78 for each participant, re-trained the model, and then made out-of-sample predictions. The predictive accuracy is good (0.828) and the confusion matrix below shows a satisfactory balance between precision and recall:

Table 5: Message Ratings Model Confusion Matrix

	Actual 0	Actual 1
Predicted 0	11	3
Predicted 1	2	13

## Sensitivity Analysis

For modeling the number of steps taken, I fit a model where I incorporated the subject day number as a fixed effect as well as interactions of the subject day number with age and gender, respectively. For this model, I used only data rows corresponding to days where a message was received by the participant. None of the additional terms were significant predictors of number of steps taken, and self-efficacy remained a significant predictor (see output in the Appendix). This suggests that there is not much of a change in the effect of the messages over the course of the study. I also fit yet another model with a delayed self-efficacy predictor to see if self-efficacy from the previous day predicted number of steps taken, but it was not a significant predictor (see output in the Appendix). Once again, the same predictors remained significant and had comparable effects to those observed in the main model.

For modeling message ratings, rather than reducing the response variable down to a binary above/below average indicator, a more natural model fit would let us keep multiple categories of the response variable. In order to see if this changed the results, I fit a hierarchical proportional-odds ordinal regression model. This model tries to fit “thresholds” between the different ordinal response categories, and a key model assumption is that these thresholds do not change with different values of the predictor variables. To meet this proportional-odds assumption, I tried multiple splits of the response variable into binned categories and chose one that met this assumption (assumption check and model output shown in Appendix). A split of the response variable that worked well was 0-6, 7-9, and 10, although using 0-5 and 6-9 for the first two splits works as well. I also had to change age from a continuous predictor to a binary indicator of whether the participant was over the age of 50 in order to meet the model assumption. The model equation is as follows:

$$\text{logit}(\gamma_{ijk}) = \log\left(\frac{\mathbb{P}(Y_i \leq j)}{1 - \mathbb{P}(Y_i \leq j)}\right)$$

where  $\gamma_{ijk} = F(\eta_{ijk})$ ,  $F$  is the inverse link function (in this case, the inverse-logit), and

$$\eta_{ijk} = \theta_k - b_j - \beta_1 \text{male}_j - \beta_2 \text{ageOver50}_j - \beta_3 \text{positive}_{ij} - \beta_4 \text{social}_{ij} - \beta_5 \text{positive:social}_{ij}$$

$j$  is the subject index from 1 to 36, and  $i$  is the observation index for a specific subject (starting at 1 and usually ending between 90 and 100, depending on the number of observations for the subject), and  $k$  is the response group index 1, 2, or 3.



Table 6: Coefficients for Prop-Odds Message Ratings Model

	Estimate	Std. Error	95% CI Lower	95% CI Upper
male	-0.4650	0.3784	-1.2219	0.2919
over age 50	-0.1412	0.3784	-0.8979	0.6155
positive	1.0708	0.1202	0.8304	1.3113
social	0.0833	0.1226	-0.1618	0.3284
positive:social	-0.3893	0.1683	-0.7258	-0.0528

From this model, we again see that compared to the baseline of negative (nonsocial) messages, positive messages predicted higher message ratings, while social messages showed no significant difference for negative messages but had a negative impact for positive messages. Age and gender also had no effect.

The ordinal package provides a built-in function to test whether or not the “proportional odds” assumption, or the assumption that the thresholds do not change with the predictor variables, is satisfied. However, the test does not work for hierarchical proportional odds models. Instead, I removed the random intercept for subject in order to run the test. When doing so the proportional odds assumption was satisfied (see Appendix). This is not an uncommon way to test the assumption in this case. [3]

As an additional sanity check, we see (in the Appendix) that removing the random intercept did not change the coefficients of the model too much, but it did change the standard errors and significance of the demographic variables, which are significant in the fixed effects model but not in the mixed-effects model. This makes sense because the variation that appeared to be between genders and ages in the fixed effects model was attributable to differences between subjects.

Overall, the results for this model were comparable to the logistic model in terms of significant predictors and their effects. In spite of its more natural fit to the data, the proportional-odds model is harder to interpret than logistic regression, so I did not choose it as my main model.

## Discussion

My goal was to use the data from this study in the Duke MCAB Lab to gain understanding about the relevance of health messages to adults and the effects of health messaging and self-efficacy on physical activity. My primary focus was the effect of the type of message received on self-reported message relevance and number of steps taken per day. I also wanted to see if there was a discrepancy between the two. Recent work in the Psychology/Neuroscience community has shown that adults are often more motivated by social rewards and positively-framed messages compared to negatively-framed messages or messages without a social component (Mikels et al. 2016), so I was also interested in whether or not these phenomena would be observed in this dataset.

I found that the adult participants in this study were more likely to give positive health messages an above-average relevance rating compared to negative health messages, consistent with prior findings. However, the effect of social components to messages was small; if anything, it seems that social components reduce the likelihood of a positive message receiving a high relevance rating. In terms of number of steps taken by participants, receiving a message did not seem to motivate participants to take more steps; in fact, the opposite was true. On the other hand, higher self-efficacy ratings seem to result in a participant taking more steps. Additionally, there is a notable

amount of heterogeneity between different participants when it comes to both steps taken and message ratings (see random intercepts for both models in the Appendix), and both measures are somewhat correlated for each participant (intraclass correlation of 0.45 for steps and 0.22 for message ratings).

The data collection process for this study introduces some limitations that are worth noting. Study participants are instructed to wear their FitBit at almost all times (except when doing things like bathing), but we cannot be sure if participants actually did this. So, if a participant's data indicates that they were unusually inactive on a given day, it could be the case that they just forgot to wear their FitBit for part of that day. Since we have no way of knowing if this was the case, I uniformly assumed that participants' daily step data during the duration of the study is accurate. Furthermore, the data that we have is only at the day-by-day granularity. The FitBit separates days at midnight, so the steps taken in one day correspond to steps taken within 24 hours of midnight when the calendar changes to that day. We have no way of knowing if this is a reasonable split point or if some participants are being very active after midnight and having this data tacked on to the next day. However, the average age of study participants is around 47-48 so I would guess that most of their walking occurs before midnight. Either way, this splitting is handled in a uniform manner by the FitBit itself and so I have no further breakdown of the data by hour in a way that would allow me to handle this.

Even with these limitations, the major takeaways from this analysis are that while adults may find positive health-related messages to be more personally relevant (as found in prior studies), these messages are not necessarily effective at motivating lifestyle changes with respect to physical activity. In contrast to prior work, I did not observe that social components to messages had a positive effect on message relevance; rather, the effect seemed to be negligible or even negative for messages that are positively-framed. Somewhat unsurprisingly, participants had notably different tendencies when it came to both steps taken and message ratings. While health messages did not motivate an increase in physical activity in this study as it was perhaps expected, there are still interesting directions for future research in this vein. For example, more studies can be done with regards to the relationship between self-efficacy and physical activity, perhaps using messages that are meant to boost self-efficacy in an effort to promote mental well-being and, by extension, physical activity.

## Citations

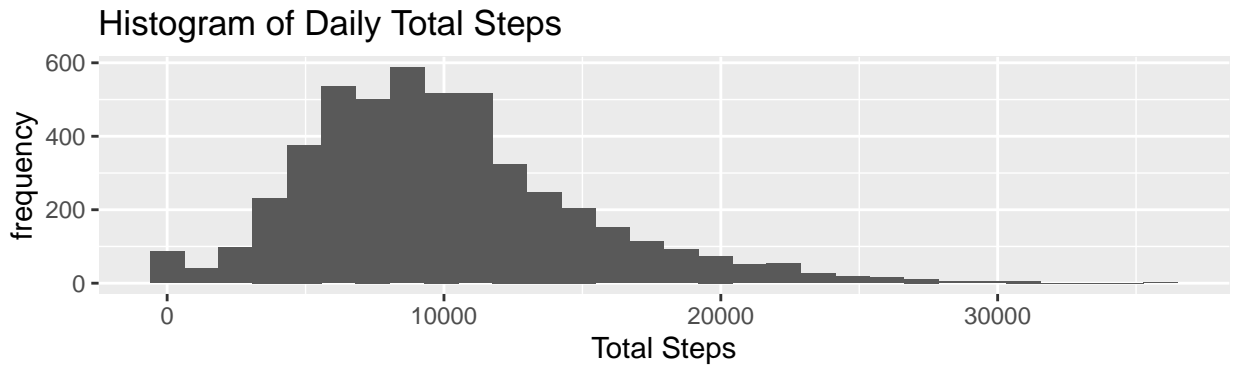
- [1] Gerber, Ben S et al. “Mobile Phone Text Messaging to Promote Healthy Behaviors and Weight Loss Maintenance: A Feasibility Study.” *Health Informatics Journal* vol. 15,1 (2009): 17-25. doi: 10.1177/1460458208099865
- [2] Mikels, Joseph A et al. “Messages That Matter: Age Differences in Affective Responses to Framed Health Messages.” *Psychology and Aging* vol. 31,4 (2016): 409-14. doi:10.1037/pag0000040
- [3] Mangiafico, S.S. “Summary and Analysis of Extension Program Evaluation in R.” version 1.18.1. (2016): [rcompanion.org/handbook/](http://rcompanion.org/handbook/)
- [4] Kraus, William E., et al. “Daily Step Counts for Measuring Physical Activity Exposure and Its Relation to Health.” *Medicine & Science in Sports & Exercise*, vol. 51, no. 6 (2019): pp. 1206–1212. doi:10.1249/mss.0000000000001932.
- [5] Pillay, Julian D, et al. “The Association between Daily Steps and Health, and the Mediating Role of Body Composition: a Pedometer-Based, Cross-Sectional Study in an Employed South African Population.” *BMC Public Health* vol. 15, no. 1 (2015). doi:10.1186/s12889-015-1381-6.
- [6] Chaumeton, Nigel, et al. “A Measurement Model of Youth Physical Activity Using Pedometer and Self, Parent, and Peer Reports.” *International Journal of Behavioral Medicine*, vol. 18, no. 3 (2010): pp. 209–215. doi:10.1007/s12529-010-9118-5.

## Appendix

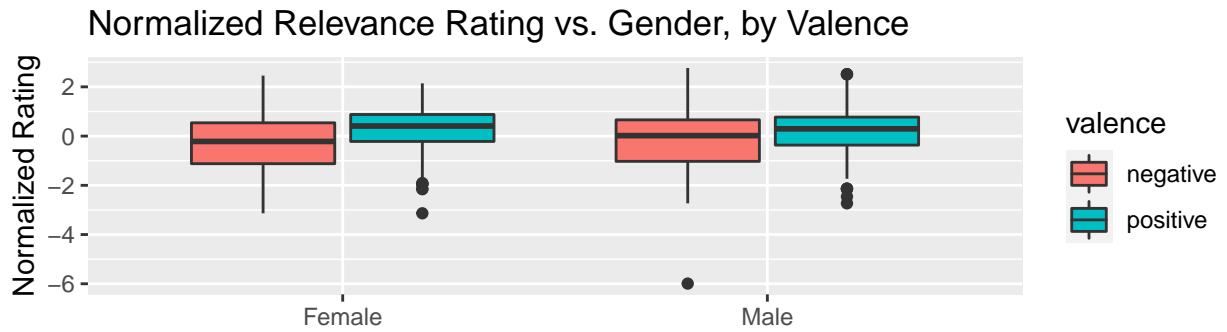
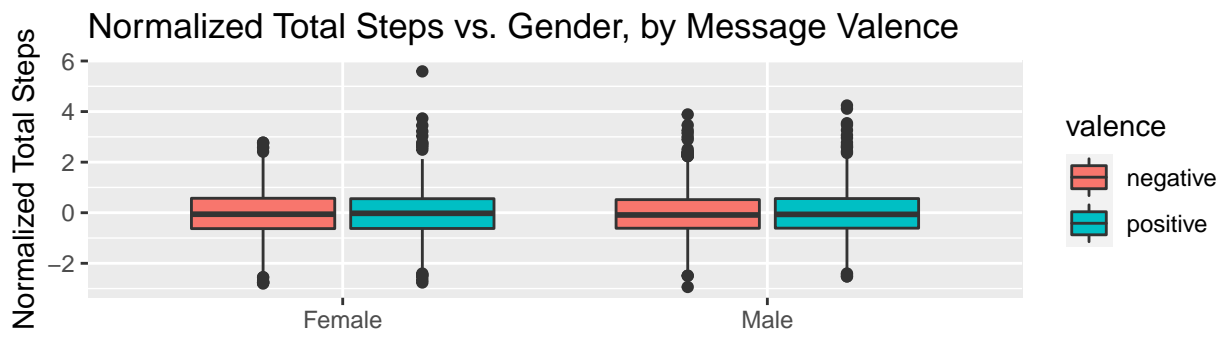
### Age and Gender T-Test

```
##  
## Welch Two Sample t-test  
##  
## data: demographics_by_sub$age by demographics_by_sub$gender  
## t = 0.90429, df = 26.909, p-value = 0.3739  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -6.854493 17.654493  
## sample estimates:  
## mean in group Female mean in group Male  
## 49.65 44.25
```

### Histograms of Response Variables

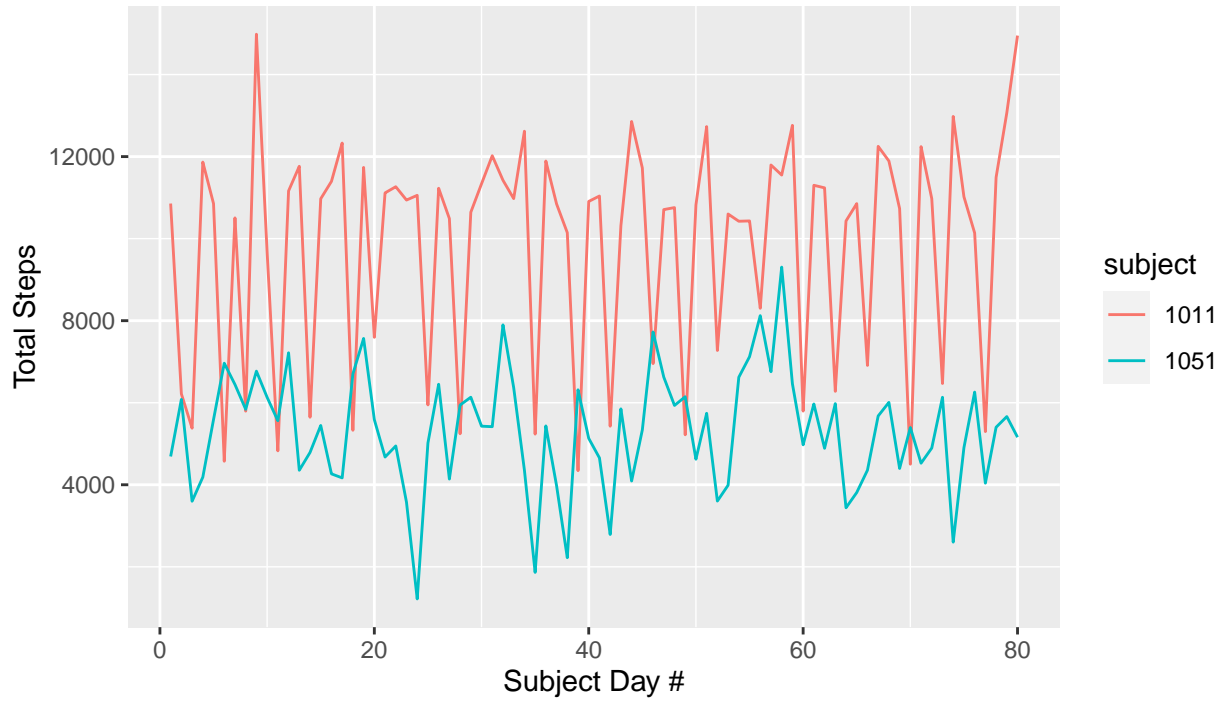


## Normalized Boxplots

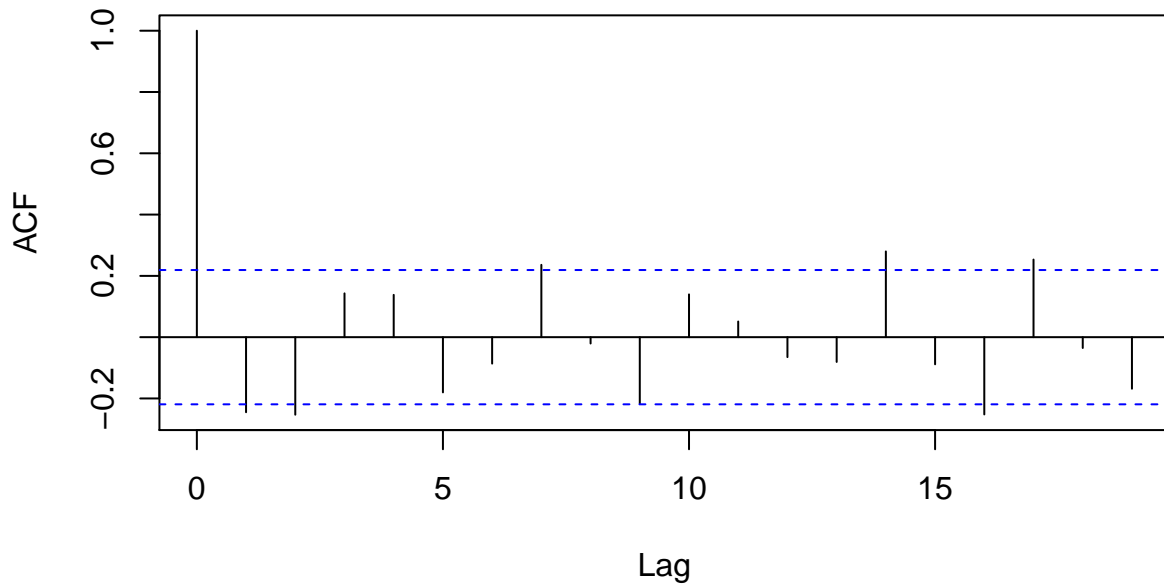


## Total Steps Autocorrelation Plots

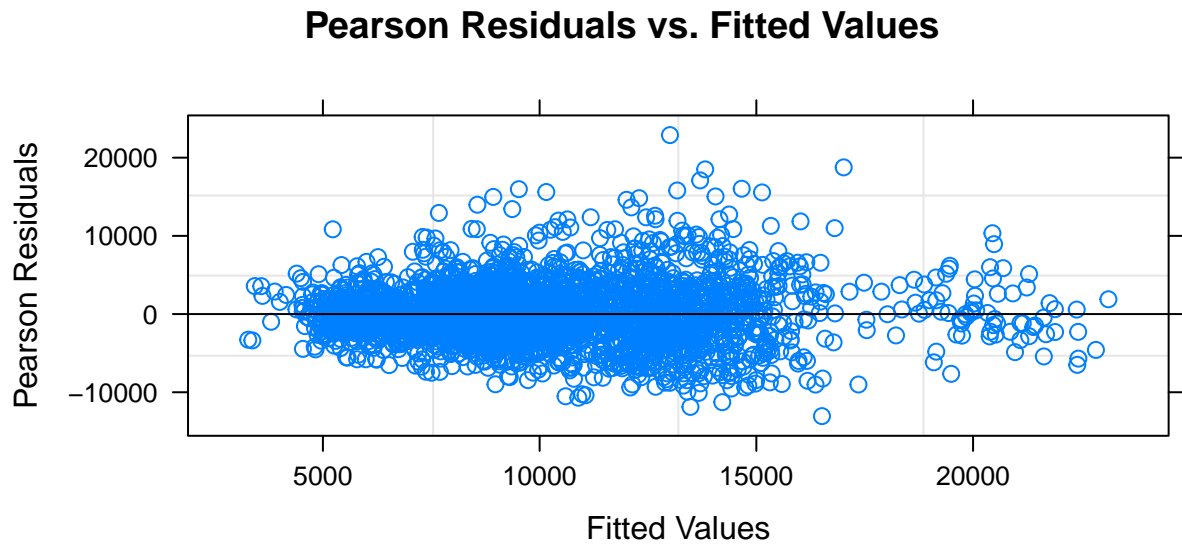
### Total Steps Over Time for Subjects 1011 and 1051



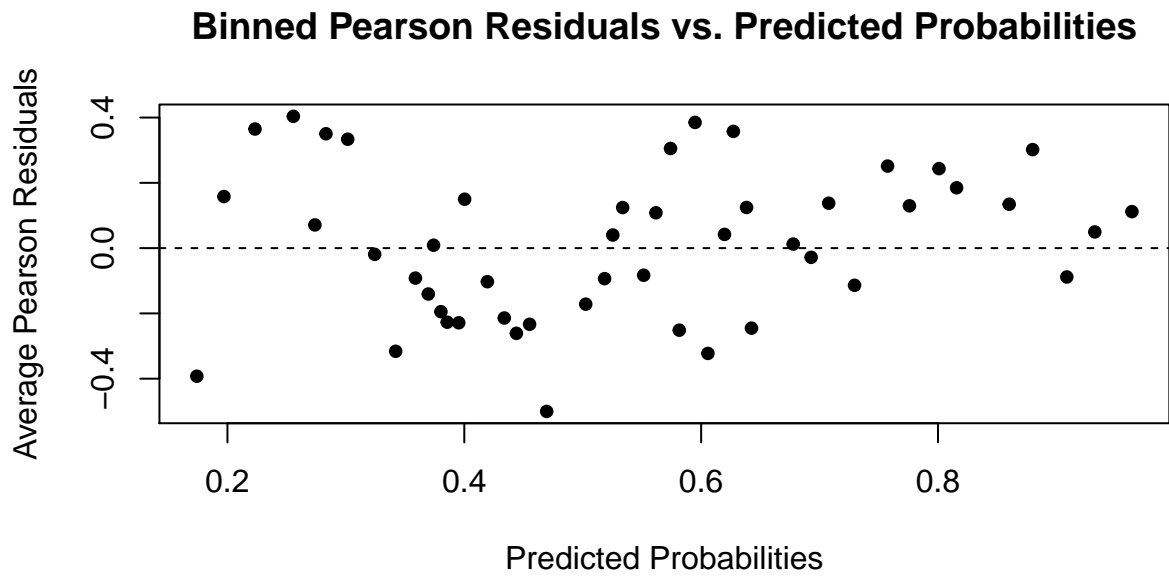
### Total Steps Autocorrelation for Subject 1011



Total Steps Main Model Residuals



Relevance Rating Main Model Residuals



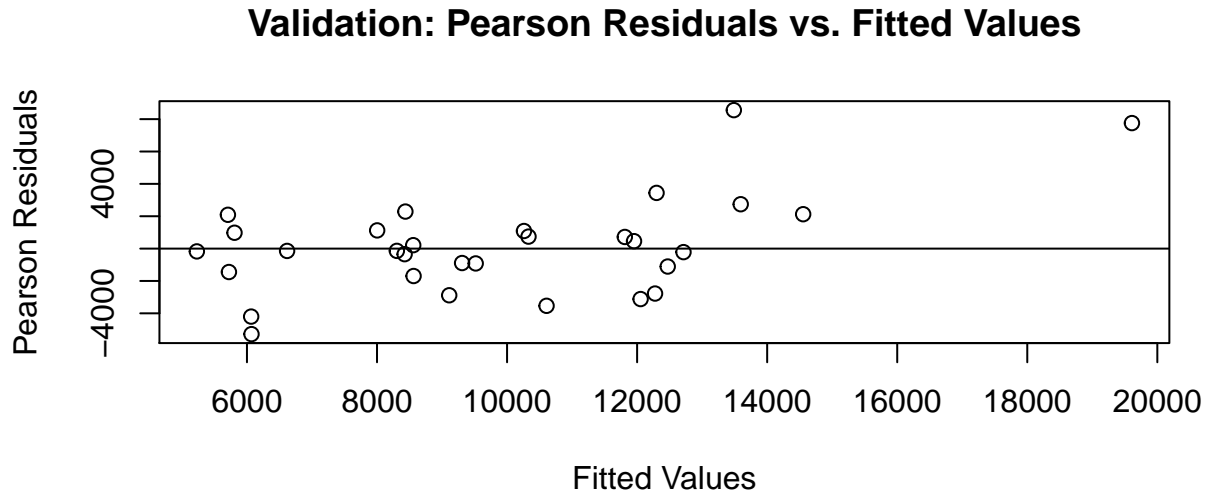
## Random Intercepts for Main Models

The table below shows the random intercepts for each participant in the study for the main steps model and the main message ratings model. For both models, these intercepts include the global intercept coefficient  $\beta_0$  added in.

	Steps Intercept	Ratings Intercept
1007	4016.691	1.4059
1010	6571.528	-1.7690
1011	8640.718	-1.2611
1019	5036.493	-0.6665
1023	5336.910	0.5904
1027	6083.050	-0.5971
1039	7570.798	-1.0520
1040	10722.236	2.5112
1044	6652.676	-0.7557
1046	11241.510	-0.4934
1048	5727.779	-0.7477
1051	2002.164	-1.0122
1053	3291.502	0.2809
1055	9160.528	-0.2201
1057	17922.251	0.6265
1058	10594.944	-0.9039
1067	9431.051	-0.2955
1071	5838.854	0.8019
1072	5217.168	-0.6320
1084	10964.153	-0.6821
1085	7051.907	-0.5147
1086	6111.751	-0.2660
1089	7815.653	0.3709
1091	3013.706	-0.0618
1093	7783.291	-0.3746
1094	10946.071	-0.3009
1095	7859.261	1.0413
1098	8717.269	0.5876
1099	12332.287	-0.3320
1101	7171.050	0.6617
1105	5562.070	-0.9179
1113	11296.441	-0.1490
1114	5456.807	1.9811



**Total Steps Validation Residuals**



**Total Steps Mixed Model with Day Number and Interactions**

	Estimate	Std. Error	95% CI Lower	95% CI Upper
(Intercept)	7723.2899522	816.5389880	6090.2119761	9356.3679283
male	-33.6716961	1192.8308478	-2419.3333917	2351.9899996
age	6.7106874	35.0397139	-63.3687403	76.7901151
self efficacy	399.5018419	37.1543855	325.1930708	473.8106130
subject day #	-4.0180680	4.2769662	-12.5720004	4.5358643
male:subject day #	2.4679990	6.4816749	-10.4953509	15.4313489
age:subject day #	-0.3339648	0.1935408	-0.7210465	0.0531168

**Total Steps Mixed Model with Self-Efficacy Lag**

	Estimate	Std. Error	95% CI Lower	95% CI Upper
(Intercept)	7716.4645351	776.7850486	6162.8944380	9270.0346323
male	40.3896186	1119.3570151	-2198.3244116	2279.1036488
age	-0.9027598	32.8398992	-66.5825582	64.7770386
self efficacy	441.5450844	33.5343209	374.4764426	508.6137263
negative nonsocial	-299.6942313	217.1005850	-733.8954013	134.5069386
negative social	-468.5534875	217.9505459	-904.4545794	-32.6523957
positive nonsocial	-522.3429748	217.6333079	-957.6095906	-87.0763589
positive social	-218.0430686	218.5010769	-655.0452224	218.9590852
1-day lagged steps	0.1414356	0.0186984	0.1040388	0.1788324
2-day lagged steps	0.0809297	0.0180614	0.0448069	0.1170525
1-day lagged self efficacy	-62.6166945	34.5814082	-131.7795110	6.5461220

## Relevance Rating Linear Mixed Model

	Estimate	Std. Error	95% CI Lower	95% CI Upper
(Intercept)	5.9762958	0.2463803	5.4835351	6.4690565
male	-0.3734937	0.3501307	-1.0737551	0.3267678
age	0.0111843	0.0104080	-0.0096316	0.0320002
positive	1.3524356	0.1122864	1.1278629	1.5770083
social	0.1876757	0.1123652	-0.0370547	0.4124061
positive:social	-0.4174038	0.1590131	-0.7354301	-0.0993775

## Relevance Rating Fixed Effects Prop-Odds Model for Checking Assumptions

```
## formula: rating_group ~ gender + over50 + valence + s_ns + valence * s_ns
## data: data_test
##
## link threshold nobs logLik AIC niter max.grad cond.H
## logit flexible 2371 -2170.80 4355.60 5(0) 1.11e-07 6.1e+01
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## genderMale -0.35100 0.08174 -4.294 1.75e-05 ***
## over501 -0.19720 0.08186 -2.409 0.0160 *
## valencepositive 0.92972 0.11402 8.154 3.52e-16 ***
## s_nsocial 0.09072 0.11432 0.794 0.4275
## valencepositive:s_nsocial -0.32602 0.16005 -2.037 0.0417 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Threshold coefficients:
## Estimate Std. Error z value
## (-1,6] |(6,9] 0.002186 0.098575 0.022
## (6,9] |(9,10] 2.515477 0.115513 21.777

## Tests of nominal effects
##
## formula: rating_group ~ gender + over50 + valence + s_ns + valence * s_ns
## Df logLik AIC LRT Pr(>Chi)
## <none> -2170.8 4355.6
## gender 1 -2169.4 4354.8 2.77567 0.09571 .
## over50 1 -2169.9 4355.9 1.74444 0.18658
## valence 1 -2170.8 4357.5 0.05346 0.81715
## s_ns 1 -2170.7 4357.5 0.14524 0.70313
## valence:s_ns 3 -2170.6 4361.1 0.49956 0.91899
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```