Getting Moving Together; a field experiment on workplace physical activity¹

Michael Sanders PhD¹², Karen Tindall PhD³, Alex Gyani PhD⁴, Susannah Hume MSc⁵,
Min-Taec Kim MSc⁶, Paul Vilanti BA⁷, Sarah Coghlan BA⁸, Anna Flego BA⁹

3843 Words

¹ Policy Institute at King's College London, 22 Kingsway, London, WC2B 6LE email:
Michael.t.sanders@kcl.ac.uk

Getting Moving Together; a field experiment on workplace physical activity¹

Question: Can social comparison 'nudges' increase the effectiveness of wearable activity trackers?

Findings: In this trial involving 646 adults, physical activity measured by number of steps per day was increased among participants who received feedback on their team's performance, relative to peer teams in the company, by 8%.

Meaning: Social comparison can motivate additional physical activity, particularly for women, the less active, and those who do not 'win' in social comparisons.

Importance: Wearable devices are widely used in an effort to increase physical activity and consequently to improve health. The evidence for this is patchy, and it does not appear that wearables alone are sufficient to achieve this end.

Objective: To determine whether social comparisons in a workplace setting can increase the effectiveness of wearables at promoting physical activity.

Design: A four week randomized controlled trial conducted in November 2015 with employees of a large firm. Participants were randomised to one

of two treatment conditions (control vs social comparison) at team level, and teams are formed into 'leagues' based on their activity levels before the study. Impact is measured through wearable devices issued to all participants throughout the study duration.

Setting: Offices of a large Australian employer.

Participants: 646 employees of an Australian employer, issued with wearable activity trackers prior to the beginning of the study.

Intervention(s) (for clinical trials) or Exposure(s) (for observational studies). Participants used a wearable device to track steps. Participants had been wearing these for at least four weeks at the outset of the trial, establishing a baseline level of activity. Teams (n=646, k=49), were randomly assigned to either control (k=24), or a social comparison (k=25) treatment.

All participants took part in a step-count competition between their team and others at their employer, in which their team's ranking within a mini-league of five teams, as well as their own activity was communicated each week. The control group had access to the usual features of the wearable, while the social comparison group received additional information about the performance of the other teams in their league, including how far behind and ahead their nearest rival teams were.

Main Outcome(s) and Measure(s): Number of steps taken per day on average, measured by the wearable devices issued to all participants.

Results: A total of 646 participants were included in the study. Compared to the control, participants in the social comparison group took

significantly more steps per day during the trial period (an additional 620 steps, 8.2%, $p < 0.001$). These effects are largest in both relative and absolute terms for people whose prior steps were in the bottom quartile of steps (an additional 948 steps, 40%, $p < 0.001$), while the effect on people with highest levels of activity was a precisely estimated null (an additional 6 steps, 0.01%, $p = 0.98$).

Conclusions and Relevance: Social comparison increased the effectiveness of wearables at improving physical activity, particularly for those with the lowest baseline activity.

Introduction

Employers are often and increasingly concerned by the wellbeing of their staff, including their physical fitness and health in a time when lifestyle disease rates are high and rising in much of the developed world. Two strategies often used by firms to support this are workplace fitness drives and competitions, and providing their staff with wearable fitness devices (or at least encouraging their use).

Physical Activity

Physical activity is an essential part of maintaining a healthy lifestyle and together with diet constitutes an essential part of avoiding lifestyle diseases such as heart disease and Type II diabetes. However, there can be substantial barriers to engaging in physical activity. One challenge is that the costs, in terms of time and effort, of this activity must be experienced in the present while the health benefits are accrued in the future. The benefits of exercise are not linear in the *amount* of physical activity undertaken and in fact experience diminishing marginal returns, so that workplaces that have a largely sedentary workforce (offices) can get more benefit by motivating people with very low levels of activity to move a little bit, than could be achieved by getting people who are already fairly active to increase their activity by the same amount.

Activity Trackers

Physical activity trackers, typically either in the form of a small device worn on the person of the user that tracks activity and its intensity through a combination of measuring acceleration of the device itself and the heart rate of the user, or in the form of an application on a smart mobile phone, have grown in use in recent years as a tool for helping people increase their fitness. The evidence on the effectiveness of activity tracking devices by themselves is mixed, with some studies^{10 11} finding either no effects or inclusive results, while some find reductions in weight loss as a consequence¹². Patel et al (2015), argue convincingly that there is a gulf between measurement and behaviour change that wearable devices are failing to bridge, and that these devices can at best act as facilitators, or delivery vectors, for other interventions¹³, rather than being interventions themselves per se. There has been relatively more success within that paradigm, with both individual and team incentives¹⁴ being shown to be effective.

Social Comparisons

That people respond to social information - information about others' behaviour, often from a relevant reference group - is a finding from behavioural science which has been widely applied in practical settings¹⁵.

In recent years, the most common formulation of social comparison has focused on those not complying with a social norm: for a given behaviour, informing people of the prevalence of a particular behaviour with the aim of increasing the proportion of people engaging in it. Providing norm information has increased, for example, tax collection, by saying that "9 out of 10 people have paid their tax"¹⁶, and environmental behaviour¹⁷.

The same form of interventions have been applied to influence the level of a behaviour among those already engaged in it by providing information on the current level of that

behaviour among other people; for example, influencing the number of doses of antibiotics doctors prescribe¹⁸ or school attendance levels⁹. Social comparisons have been found to be effective when compared to a control group and financial incentives in the field of physical activity²⁰.

These social comparisons may work through a combination of channels: by exploiting people's desire to conform to group behaviour²¹, or to compete with their peers, coupled with either inaccurate or incomplete beliefs about others' behaviour, or a lack of salience of others' behaviour.

The social distance between the participants and the comparison group is also relevant. Hallsworth et al (2017) find that in tax compliance, when the comparison group is the entire country, the same intervention is less effective than when the comparison group is people within the recipient's town. The workplace, in which people spend the largest chunk of their waking hours²², could offer fertile ground for social comparison

There are, however, situations in which social comparisons may not be helpful, for instance where the behaviour of interest is not very prevalent, or where competition or conformity is already occurring. There are also groups for whom we would posit it would not be effective, particularly those who are already performing at or above the comparison point; for instance, studies in energy efficiency find either no effect²³ or a 'boomerang' effect when low energy users are informed their consumption is better than average. For better-than-average performers, the desire to conform to the average, or a sense of shame at underperformance, would not motivate a change in behaviour.

Material and Methods

Our experiment was carried out in a large nationwide company in the construction sector in Australia. Participants were employees of the company who had been offered subsidised wearable activity trackers by their employer, who had taken part in a "Fitbit challenge" the previous month, in which teams were encouraged to aim to take 10,000 steps per day, recorded by their wearable device, and could win small prizes as well as recognition inside the company.

Teams are structured in the same way as the company, which follows a format of *location - function*, for example *Queensland - Accounts*. The teams in which people competed during the challenge were therefore teams of close colleagues who were largely located in the same physical space.

At the conclusion of the first Fitbit Challenge, in October 2015, participants were informed that the company was running a "rematch", to continue the success of the previous month. Neither the original challenge nor the rematch offered incentives or prizes for high performing teams. At this point, leagues of teams, called "Challenge Groups" by the company, were formed based on their performance in the first challenge. The first Challenge Group contained the teams with the 1st to the 5th best performance in the original challenge, the second group contained the 6th-10th best performing groups, and so on. Challenge groups were formed of five teams in order to provide each team with a comparison set that

was likely to be similar in performance to themselves, and to avoid either the best team growing complacent or lower performing teams being disheartened by being set a comparison that was too remote. Because of the way that Challenge Groups were formed, it is possible for the teams within a Challenge Group to be highly geographically and functionally separated from each other.

The sample for the experiment consisted of 646 individuals, working in 50 teams that were been divided into ten Challenge Groups. Of the participants, 333 (51.4%) were female. In the period prior to the experiment, the average number of steps per person/day was 7905, with a minimum of 0 and a maximum of 37,494. Teams consisted on average of 13 participants, with the smallest team containing two participants and the largest 62.

Teams were randomised into one of two conditions. All participants received weekly emails or mobile phone push notifications via the Fitbit mobile phone app about the challenge and how it was progressing, company wide information about physical activity, and generic information on the performance of their own team. Whether the message was received by email or push notification depended on the user's notification setting in the Fitbit app.

Generic Messages (Control) condition: participants in this condition received only the general information above.

Social Comparison (Treatment) condition: participants in this condition received the general information; but in addition were told which team in their Challenge Group was in the lead, and how many more steps they would need to take in order to beat them. An example of the additional information given in this condition is given below.

"Your team is 2nd in your challenge group with an average of 12,905 steps...your team is only 1,070 behind Queensland - Accounts, who is leading your challenge group."

Notifications of this form were repeated each week, with updated information for the challenge to date.

Balance checks conducted at the beginning of the study show no significant imbalance on pre-treatment average number of steps taken per day ($p=0.65$), the primary outcome measure, or on either participants' pre-treatment minutes of exercise ($p=0.51$), or the number of days in which they achieve their 10,000 step target ($p=0.372$).

Main Results

Our main outcome measure is the number of steps taken by participants on an average day in the week. We collapse noisy daily steps data to produce weekly data for the duration of the trial. Also included in our analysis are pre-treatment data gathered during the initial Fitbit Challenge. These data are included in order to increase the level of precision of our estimates given the relatively modest sample size.

Reflecting the structure of our data, the analysis uses a panel regression with participant level fixed effects, and clustering at the level of the team. The results of regressions on our primary outcome measure - number of steps taken per day, averaged over a week - are reported in Table 1, below. We see a positive and significant ($p < 0.001$), increase in the number of steps taken as a result of our treatment, amounting to an additional 620 steps per day on average, and a gradual decline in the number of steps taken per day over the course of the study, with 243 steps fewer being taken per day on average in each passing week. The average number of steps taken over the experimental period is 7528 per person per day, and so our effect amounts to an 8.2% increase. As described above, our data contain some people with a very large number of steps, and some with a very low number of steps (0). e There is no reason to anticipate that any measurement error is correlated with treatment, and so present our intention to treat analysis as primary. However, we do not find substantive changes to our findings where participants with 0 steps are omitted.

Table 1: Main effects on number of steps,, experiment 1
(1)
Primary Analysis
(Steps)

Treatment	620.117*** (132.914)
Week	-243.132*** (20.135)
Constant	8430.579*** (63.056)
Experimental Period Mean	7528
<i>N</i>	4448

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, fixed effects panel regression estimates with standard errors clustered at the level of the participants' team.

Secondary Analyses

As has been described above, there is an open question regarding for whom social comparisons are effective at inducing changes in behaviour. In our secondary analysis we analyse the results for participants with different levels of physical activity in the baseline period, split by quartile of prior exercise.

The results of these analyses are shown in Table 2, below. We find that effects are largest, both in relative and absolute terms, for participants with low prior step counts, with the bottom quartile increasing their steps by 945 ($p < 0.001$), a relative increase of 40%. By comparison, participants in the top quartile of prior steps see a statistically insignificant increase of just 6 steps ($p = 0.98$).

Table 2: Effects of social comparison on exercise, by prior exercise levels.

	(1) (1st quartile)	(2) (2nd quartile)	(3) (3rd quartile)	(4) (4th quartile)
Treatment	945.794*** (259.140)	1046.596*** (281.063)	615.304* (250.587)	6.429 (266.253)

Week	-95.768** (36.957)	-387.353*** (44.701)	-271.338*** (38.697)	-229.856*** (39.763)
Constant	2207.344*** (117.455)	7561.996*** (137.990)	9923.761*** (120.119)	14020.937*** (125.507)
<i>Experimental Period Mean</i>	2330	6264	8774	12842
<i>N</i>	1111	1110	1112	1115

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, fixed effects panel regression estimates with standard errors clustered at the level of the participants' team.

Finally, we can look at teams' positions within their own Challenge Groups at the outset of the trial - that is, their performance relative to their competition during the pre-experiment period. Because the team that is 5th in the 1st challenge group starts in "last place", and the team that is in 6th, by virtue of being in the 2nd challenge group starts in "first place" - despite having been less active in the pre-experiment period, we are able to estimate the effect of these ordinal rankings on subsequent performance, shown below in Table 3. This, novelly, allows us to investigate the difference between absolute performance (how many steps a participant took), and relative performance (rank), in responsiveness to social information.

Table 3: Exploratory Results by initial position in challenge group - effects of treatment on total steps

	(1) (1st/2nd)	(2) (3rd)	(3) (4th/5th)
Treatment	379.486* (187.597)	956.995*** (271.938)	753.509** (291.592)

Week	-213.933*** (32.531)	-279.260*** (45.020)	-250.861*** (31.135)
Constant	8213.111*** (96.531)	9147.770*** (135.820)	8217.475*** (105.462)
<i>N</i>	1863	1023	1562

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, fixed effects panel regression estimates with standard errors clustered at the level of the participants' team.

Here, we find significant effects for all groups, regardless of their starting position, but that the sizes of the treatment effects vary quite substantially by baseline position, with a 956 step increase observed for participants who start in a third place team, compared to a 379 step increase for participants who start in a 1st or 2nd placed team, and these differences are significantly different ($p < 0.05$). Teams that begin in 4th or 5th place see an increase of 759 steps, falling between the other two groups, and are not significantly different from the others.

Discussion

We test the impacts of social comparison messaging on physical activity in a workplace competition. In common with the previous literature, we find that information on the performance of peers is effective at increasing steps. Similarly, we find that this is concentrated on participants who are at the bottom or middle of the distribution of physical activity, with no effect for the most active participants. Exploiting that rankings provided in our treatment are not perfectly correlated with actual physical activity as they otherwise might be, due to our Challenge Group structure, we find, again in keeping with the goal-setting literature, that participants are most motivated by social information which suggests that they do not have too far to go in order to overtake the leaders, compared to being told that they are in 4th or 5th place out of five.

Taken together, these findings suggest that social comparison can motivate many participants, but that high performing participants, either in absolute or relative terms, are likely to be harder to motivate using this approach.

References

Reiner, M., Niermann, C., Jekauc, D., & Woll, A. (2013). Long-term health benefits of physical activity—a systematic review of longitudinal studies. *BMC public health*, 13(1), 1-9.

John P, Sanders M, Wang J. The use of descriptive norms in public administration: a panacea for improving citizen behaviours? 2014 Available at SSRN: <https://ssrn.com/abstract=2514536> or <http://dx.doi.org/10.2139/ssrn.2514536>.

Shuger SL, Barry VW, Sui X, et al. Electronic feedback in a diet-and physical activity-based lifestyle intervention for weight loss: a randomized controlled trial. *International Journal of Behavioral Nutrition and Physical Activity*. 2011; 1;8(1):41.

Harris T, Kerry SM, Victor CR, et al. A primary care nurse-delivered walking intervention in older adults: PACE (pedometer accelerometer consultation evaluation)-Lift cluster randomised controlled trial. *PLoS Med*. 2015; 17;12(2):e1001783.

Jakicic JM, Davis KK, Rogers RJ, et al. Effect of wearable technology combined with a lifestyle intervention on long-term weight loss: the IDEA randomized clinical trial. *JAMA*. 2016; 20;316(11):1161-71.

Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. *JAMA*. 2015; 3;313(5):459-60.

Patel MS, Asch DA, Rosin R, et al. Individual versus team-based financial incentives to increase physical activity: a randomized, controlled trial. *Journal of General Internal Medicine*. 2016; 1;31(7):746-54..

Hallsworth M, List JA, Metcalfe RD, Vlaev I. The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*. 2017; 1;148:14-31..

Goldstein NJ, Cialdini RB, Griskevicius V. A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*. 2008; 1;35(3):472-82..

Hallsworth M, Chadborn T, Sallis A, et al. Provision of social norm feedback to high prescribers of antibiotics in general practice: a pragmatic national randomised controlled trial. *The Lancet*. 2016; 23;387(10029):1743-52.

Rogers T, Feller A. Reducing student absences at scale by targeting parents' misbeliefs. *Nature Human Behaviour*. 2018; 2(5):335-42.

Patel MS, Volpp KG, Rosin R, et al. A randomized trial of social comparison feedback and financial incentives to increase physical activity. *American Journal of Health Promotion*. 2016; 30(6):416-24..

Bernheim BD. A theory of conformity. *Journal of Political Economy*. 1994; 1;102(5):841-77..

Endnotes

1. We are grateful to the Movember Foundation for funding this research. We are also grateful to Jason Wedesweiler from Lendlease Australia. We are grateful to seminar participants at the University of Pennsylvania, and participants at the Harvard Behavioural Insights and Public Health conference.
2. King's College London
3. Behavioural Insights Team and Institute for Governance & Policy Analysis, University of Canberra
4. Behavioural Insights Team
5. King's College London
6. Behavioural Insights Team
7. The Movember Foundation
8. The Movember Foundation
9. The Movember Foundation
10. Shuger SL, Barry VW, Sui X, et al. Electronic feedback in a diet-and physical activity-based lifestyle intervention for weight loss: a randomized controlled trial. *International Journal of Behavioral Nutrition and Physical Activity*. 2011; 1;8(1):41.
11. Harris T, Kerry SM, Victor CR, et al. A primary care nurse-delivered walking intervention in older adults: PACE (pedometer accelerometer consultation evaluation)-Lift cluster randomised controlled trial. *PLoS Med*. 2015; 17;12(2):e1001783.
12. Jakicic JM, Davis KK, Rogers RJ, et al. Effect of wearable technology combined with a lifestyle intervention on long-term weight loss: the IDEA randomized clinical trial. *JAMA*. 2016; 20;316(11):1161-71.
13. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. *JAMA*. 2015; 3;313(5):459-60.
14. Patel MS, Asch DA, Rosin R, et al. Individual versus team-based financial incentives to increase physical activity: a randomized, controlled trial. *Journal of General Internal Medicine*. 2016; 1;31(7):746-54..
15. John P, Sanders M, Wang J. The use of descriptive norms in public administration: a panacea for improving citizen behaviours? 2014 Available at SSRN: <https://ssrn.com/abstract=2514536> or <http://dx.doi.org/10.2139/ssrn.2514536>.
16. Hallsworth M, List JA, Metcalfe RD, Vlaev I. The behaviorist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*. 2017; 1;148:14-31..
17. Goldstein NJ, Cialdini RB, Griskevicius V. A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*. 2008; 1;35(3):472-82..
18. Hallsworth M, Chadborn T, Sallis A, et al. Provision of social norm feedback to high prescribers of antibiotics in general practice: a pragmatic national randomised controlled trial. *The Lancet*. 2016; 23;387(10029):1743-52.
19. Rogers T, Feller A. Reducing student absences at scale by targeting parents' misbeliefs. *Nature Human Behaviour*. 2018; 2(5):335-42.
20. Patel MS, Volpp KG, Rosin R, et al. A randomized trial of social comparison feedback and financial incentives to increase physical activity. *American Journal of Health Promotion*. 2016; 30(6):416-24..
21. Bernheim BD. A theory of conformity. *Journal of Political Economy*. 1994; 1;102(5):841-77..
22. Authors' analysis of the American Time Use Survey
23. Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological science*, 18(5), 429-434.