Inspiration from the "Biggest Loser": Social Interactions in a Weight Loss Program*

Kosuke Uetake† Nathan Yang‡

May 3, 2018

Abstract

We investigate the role of heterogeneous peer effects in encouraging healthy lifestyles. Our analysis revolves around one of the largest and most extensive databases about weight loss that track individual participants' meeting attendance and progress in a large national weight loss program. The main finding is that while weight loss among average performing peers has a negative effect on an individual's weight loss, the corresponding effect for the top performer among peers is positive. Furthermore, we demonstrate that our results are robust to potential issues related to selection into meetings, endogenous peer outcomes, individual unobserved heterogeneity, lagged dependent variables, and contextual effects. Ultimately, these results provide guidance about how the weight loss program should identify role models.

Keywords: Big Data; Customer Relationship Management; Healthy Living; Subscription Services; Weight Management

*Forthcoming in *Marketing Science*. We thank Yikun Jiang for excellent research assistance. We are also grateful to the Senior Editor, Associate Editor, and two anonymous reviewers for helpful suggestions throughout the review process. Finally, this project benefited from discussions with conference and seminar participants at the McGill Desautels Faculty of Management, Marketing Science Conference on Health, University of Alberta, McGill Centre for the Convergence of Health and Economics Consumer Behavior Webinar, Yale School of Management, and Yale Human Neuroscience Lab, as well as Rahi Abouk, Francesco Amodio, Neeraj Arora, David Buckeridge, Tat Chan, Hai Che, Nicholas Christakis, Victor Couture, Jean-Pierre Dubé, Laurette Dubé, Liran Einav, Gautam Gowrisankaran, Linda Hagen, Gerald Häubl, Mitchell Hoffman, Raghuram Iyengar, Sandra Laporte, Peter Popkowski Leszczyc, Yu Ma, Puneet Manchanda, Tanya Mark, Brent McFerran, Saurabh Mishra, Ashesh Mukherjee, Sridhar Narayanan, Oded Netzer, Yeşim Orhun, Margarita Pivovarova, John Pracejus, Brian Rubineau, Juan Serpa, K. Sudhir, Michael Tsang, Ishani Tewari, Kangkang Wang, Chuck Weinberg, and Sunghwan Yi.

†Yale School of Management. Email: `kosuke.uetake@yale.edu`.

‡McGill Desautels Faculty of Management, CIREQ, CIRANO, GERAD, CCHE, and MCCHE. Email: `nathan.yang3@mcgill.ca`.

1 Introduction

Healthy lifestyles are valued by customers and firms alike. In addition to direct behavioral interventions (e.g., Hagen, Krishna, and McFerran, 2016), such lifestyles may in fact propagate throughout a social network via interactions and peer effects, as indicated by past research showing peer effects in health outcomes (e.g., Christakis and Fowler, 2008). However, not all peers are alike, and thus, each peer's impact on others within the group need not be homogeneous. In light of this heterogeneity, how should firms identify role models in their efforts to promote healthy lifestyles?

Our research studies the impact of heterogeneous peer effects under the context of a large weight loss program in the United States, where social support from peers may play an important role in weight loss. The weight loss industry in the US is particularly large and generates about $20 billion each year from over 100 million dieters. From a commercial weight loss program's perspective, customer-centric program design policies aimed to shape and optimize the interactions between participants may have a positive impact on the level of engagement. Most importantly, customer satisfaction and development will likely be tied to the perceived performance of the program (e.g., Anderson, Fornell, and Lehmann, 1994; Kumar, Umashankar, Kim, and Bhagwat, 2014) as reflected by sustainable weight loss progress.

Heterogeneous peer effects have potentially large implications in the weight loss context. As weight loss can be thought of as a personal goal, there are often peers that are disproportionately more instrumental or focal in affecting the ability or motivation to reach the goal (e.g., Fitzsimons and Fishbach, 2009). For example, information disclosure about (relative) performance likely has an impact on motivation (e.g., Karlsson, Loewenstein, and Seppi, 2009; Lockwood et. al., 2005). The weight loss program then has to decide what information about peer performance should be disclosed. If we think of the weight loss program as a form of health education, then past insights in education and psychology would suggest motivation is delicate and quite sensitive to social comparison (e.g., Blanton, Buunk, Gibbons, and Kuyper, 1999; Rogers and Feller, 2016); furthermore, the fact that individuals can make either upward or downward comparisons with others would suggest the possibility that social comparisons may be *heterogeneous* (e.g., Buunk and Gibbons, 2007).

For our empirical analysis, we are able to track each participant's weight loss progress at a high frequency for more than a million users, as well as their group meeting attendance. At these meetings, participants weigh-in, interact with other weight loss participants, and consult with a weight-loss mentor. With this data, we investigate the impact that peer weight loss has on an individual's weight loss success. Based on a variant of the standard linear-in-means peer effect framework (Brock and Durlaf, 2001; Manski, 1993), we allow the peer effect to be heterogeneous across performance groups by categorizing peers at a given meeting as *best*, *average*, and *worst* performers (relative to those attending the same meeting).

A few key findings emerge. Our estimates demonstrate that the average weight loss among peers has a *negative* (i.e., discouraging) effect on an individual's own future weight loss. In contrast, we find that weight loss of the top performer has a *positive* (i.e., encouraging) effect on an individual's own future weight loss. Such findings have implications on how weight loss program employees at the meetings promote the past successes of their participants, as the successes among average participants may act as a discouraging benchmark that roughly half of the participants will fail to reach, while the successes among top performers may act as an encouraging target that does not alienate as many of the participants. Furthermore, we demonstrate that our results are robust to potential issues related to selection into meetings, endogenous peer outcomes, individual unobserved heterogeneity, lagged dependent variables, and contextual effects.[1] Based on our estimates, we later discuss implications of our findings on meeting design for the commercial weight loss program

[1] To address issues about selection and endogeneity, we make use of a combination of instrumental variables and selection correction (Olsen, 1980). The instrumental variables are constructed using detailed information about each participant's physical distance to the meeting locations along with location-specific weather patterns (i.e., temperature, precipitation). Furthermore, selection correction is implemented by including a control function.

in terms of *content* and *composition*. Regarding the content, the meeting leaders can, for example, use the weight loss successes of top performers provide inspiration to the group, and perhaps avoid using the overall group's success as the benchmark. Moreover, the weight loss program can design composition of groups of meeting participants that would maximize the encouraging effects of top performers and minimize the discouraging effects of average performers.

This study is related to past work that aims to identify heterogeneous peer effects. In particular, marketing research has demonstrated that the strength of peer effects may differ depending on the peers' spatial proximity (e.g., Bell and Song, 2007; Bollinger and Gillingham, 2012; Choi, Hui, and Bell, 2010; Gardete, 2014; Manchanda, Xie, and Youn, 2008), observable physical characteristics (e.g., McFerran, Dahl, Fitzsimons, and Morales, 2010a, 2010b; Park and Manchanda, 2014), intra-group relationship (e.g., Narayan, Rao, and Saunders, 2011; Yang, Narayan, and Assael, 2006), and level of opinion leadership or network tie strength (e.g., Aral and Walker, 2014; Godes and Mayzlin, 2009; Iyengar, Van den Bulte, and Valente, 2011; Nair, Manchanda, and Bhatia, 2010). The aforementioned work in marketing has studied effects of heterogeneous peer *actions* under the context of product, service, technology, and media consumption. In contrast, we are studying the effect of heterogeneous peer *outcomes*, in the form of their past weight loss performance. We believe our empirical context is uniquely well-suited to help us identify encouraging role models as the peer outcomes are a reflection of their weight loss abilities and efforts, which brings us to the second stream of literature that our work is related to.

In social psychology, researchers have investigated the impact of social comparisons with high-performing peers on self-evaluation. Some notable examples include Brewer and Weber (1994) and Pelham and Wachsmuth (1995). Collectively, these studies have demonstrated that top performers can either be encouraging or discouraging. For example, a top performer may help provide additional motivation to achieve similar accomplishments; but on the other hand, top performers may be demoralizing and lead individuals to think that their achievements are inadequate. Taken together, the fact that behaviorally top performers can either be encouraging or discouraging provides further justification that the impact of top performers remains an important empirical question. This past work has largely been confined to behavioral experiments, so we complement this literature by providing insights using a very large data-set from a large commercial weight loss company. Distinguishing qualities of our data-set include rich variation in meeting attendance along with the distribution of peer performance from one meeting to the next; we believe such data qualities would be difficult to achieve in a laboratory setting.

This paper proceeds with the following structure. We first provide a detailed description of the empirical setting in Section 2, along with information about key data variation in meeting composition and peer outcomes. Our empirical analysis of heterogeneous peer effects is presented in Section 3, where we first present our empirical and identification strategy, followed by a discussion of our main findings, robustness checks, and exploration of moderating effects. We then conclude in Section 4 with a brief discussion of managerial implications and future research possibilities related to our study.

2 Empirical Setting

2.1 Details About Weight Loss Program

Our analysis uses data from a large national weight loss program with nearly two million participants. The weight loss program is based in the United States, and generated about $1.7 billion in revenue during 2013. Unlike some of the other popular diet programs, the weight loss program we study does not explicitly restrict certain food groups (i.e., carbohydrates, fat, sugar, protein). Instead, they adopt a calorie budgeting system, which gives participants the freedom to eat any type of food, provided that they do not exceed their allowed calorie budget (which may increase with exercise). A unique feature of the weight loss program is that participants attend meetings and interact with other participants. Furthermore, this program's efficacy has been validated via numerous scientific and independently conducted studies.

In-person group meetings are an important component of the weight loss program. In addition to keeping track of weight loss progress, individuals have an opportunity to interact with their peers and group mentors. To get more information about the meetings, we reached out to company representatives both via phone and in person. During a call to the weight loss program's headquarters on July 19, 2016, one of the authors asked about what activities occur during each meeting, and the membership sales representative said that "sharing of experiences with other members" is the primary and critical purpose of the meeting. There may also be discussions about weekly topics, but in general, the purpose is to allow members to be inspired and hopefully benefit from the experiences and successes of others. Meetings are typically 30 minutes long, where the first 10 minutes are used by the weight loss mentor to discuss general tips about healthy lifestyles, and the remaining 20 minutes are normally allowed for social interactions. It is common to go around the room to give every participant an opportunity to share his or her weight loss experience in the past weeks or so. Virtually all participants are voluntarily transparent with one another about specifics of their weight loss progress (i.e., changes in weight, things that did and did not work for weight loss). Furthermore, there are no systematic efforts (during the period we study) to showcase the successes of certain individuals in the group. That is, a participant in the meeting is exposed to the entire distribution of peer outcomes. On November 29, 2016, one of the authors was given permission to attend a meeting, and his observations are consistent with the details provided during the conversation with the membership sales representative.

2.2 Descriptive Patterns in the Data

2.2.1 Weight Loss

To see what a typical weight loss participant looks like, Table 1 provides the summary statistics for our sample. From this table, we see that a typical weight loss participant is about 85 kilograms, 65 inches, 51 years old, and female. Note that the average weight for an American female over 20 years old is about 75 kg according to the Centers for Disease Control and Prevention (CDC). Furthermore, the average body mass index (BMI) in our sample is a bit over 31, while a healthy BMI ranges from 18.5 to 24.9. Finally, we also observe how far an individual's weight is to their long-term weight loss goal, which was determined upon joining the program.

Table 1: Summary Statistics

Variable	Mean	Std. Dev.
Weight (kg)	85.89	21.41
Height (inches)	65.16	3.16
BMI	31.27	7.04
Weight loss per day (kg)	0.04	0.13
Age	54.02	17.83
Male	0.09	0.29
Distance to goal (kg)	6.25	7.03

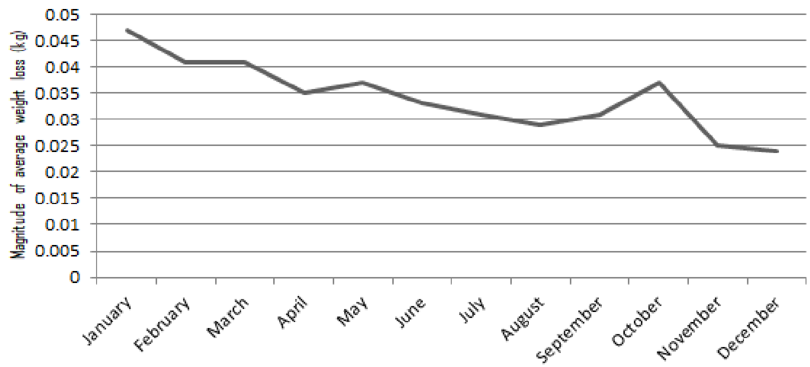
Note: Averages and standard deviations calculated across all meetings.

Since weight loss is the main outcome of interest, we provide details about weight loss dynamics. We see that from one meeting to the next, weight loss per day is on average 0.04 kilograms, which is considered to be a healthy weight loss rate according to CDC guidelines despite it seeming small. Furthermore, we can explore descriptive patterns of dynamics and seasonality in weight loss. From Figure 1, it appears that the amount of daily weight loss is largest in January and February, and trends downwards towards December. The improvement in weight loss may correspond with promotional efforts by the weight loss program around September to October.

2.2.2 Meetings

As weight loss meetings are a distinctive feature of this weight loss program, we provide some information about meeting attendance patterns (see Figure 2). Weight loss participants on average attend about 11 meetings from 2012 to 2013. Meeting locations are spread out across the United States, where there are about 1,070 official meeting locations. Individuals typically attend meetings held at the same physical location (see Panel 1). Finally, we see

Figure 1: Magnitude of Weight Loss Per Day (kg) by the Month



from Panel 2 that a large proportion of participants (about 40%) attend meetings within the same zip code as where they live. Nevertheless, there is still a large fraction of participants who travel beyond their zip code to a meeting location (e.g., traveling more than 20 km to meeting). To calculate the distance between each participant to the location of the last meeting attended, we compute geographic distances "as the crow flies" using longitude and latitude coordinates provided in the data. In terms of meeting attendance dynamics, Panel 3 shows that in over half of the observations, less than a week separated the current and previous meeting. Furthermore, in most of the observations, less than a month separates current and previous meetings. There are some spikes in the distribution as meetings are often held only on certain days of the week.

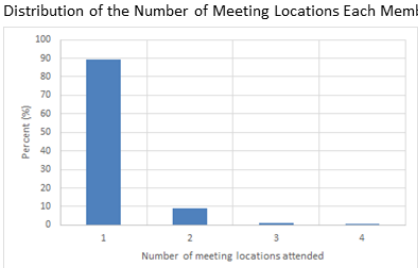
2.2.3 Composition of Peers and Peer Weight Loss Performance

Our empirical analysis is centered around understanding heterogeneous peer effects, so we will highlight variation in the composition of meeting attendees followed by a description of dynamics in peer weight loss performance (see Figure 3). From meeting to meeting, our data confirms that the composition of peers changes, sometimes drastically so.[2] Another way

[2]On average, nearly 26,000 participants will pass by a particular location during our sample, and each meeting would consist of only 0.2% of this entire pool of potential attendees. Furthermore, each location holds on average about 94 meetings per month (or about 3 meetings per day). The weight loss company offers many meeting time options to accommodate for peoples' varying schedules.

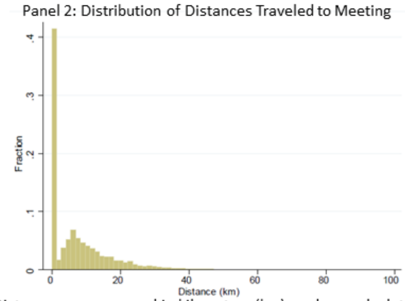
Figure 2: Descriptive Patterns in Meeting Attendance

Panel 1: Distribution of the Number of Meeting Locations Each Member Attends



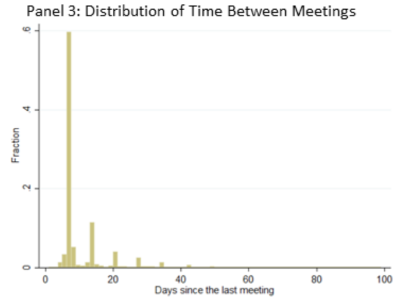
Note: Number of meeting locations calculated to be the number of unique locations a user has visited over the course of our data sample.

Panel 2: Distribution of Distances Traveled to Meeting



Note: Distances are measured in kilometers (km), and are calculated based on longitude/latitude coordinates of the participant's home address and meeting location.

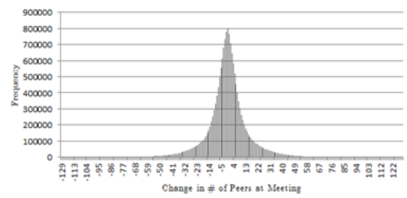
Panel 3: Distribution of Time Between Meetings



Note: Time between meetings calculated as the days that separate a current and previous meeting attended by an individual.

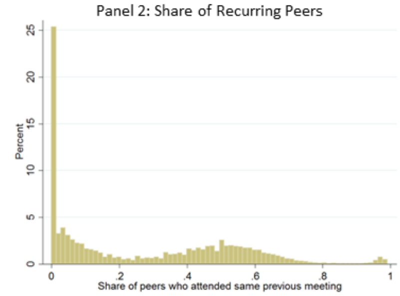
Figure 3: Descriptive Patterns in Peer Composition

Panel 1: Distribution of the Changes in the Number of Meeting Participants



Note: Change in the number of peers is calculated based on the difference between the total number of participants at current meeting versus the number of participants in the previous meeting.

Panel 2: Share of Recurring Peers



Note: The share is calculated by dividing the number of peers in current meeting who were in the same previous meeting as the focal individual, with the total number of peers in the current meeting. This histogram provides distribution of these calculated shares.

Panel 3: Dynamics in Relative Performance Rank

Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Sum across columns
1	0.09	0.06	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	102,314
2	0.07	0.06	0.05	0.05	0.05	0.05	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	123,896
3	0.07	0.06	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	138,090
4	0.06	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	147,905
5	0.06	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.04	155,091
6	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.04	0.04	160,670
7	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.04	0.04	165,869
8	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.04	0.04	169,020
9	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.04	0.04	172,922
10	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.04	0.04	174,539
11	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.04	0.04	177,101
12	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.04	0.04	177,660
13	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.04	177,933
14	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.04	176,582
15	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.04	0.04	174,141
16	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.04	170,988
17	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.04	166,017
18	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.04	160,670
19	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	153,999
20	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	146,913

Note: Rows represent current weight performance rank, while columns represent future weight performance rank. Performance rank is calculated by ordering the weight loss outcomes, from best to worst. The final column adds up the total number of observations in each cell across the columns.

to demonstrate variation in meeting group composition dynamics is to plot the histogram for the change in number of peer participants that an individual faces from one meeting to the next. Panel 1 shows us the distribution summarizing changes in peer group size. This histogram demonstrates that each individual likely faces a different set of peers across meetings. In fact, in only about 5% of the observations do we see no change in the number of peers, and in only 10% of the observations do we see a change of only 1. On a similar note, we can confirm that there is sufficient turnover from one meeting to the next such that on average, only 27% of the peers attended the same previous meeting as an individual. Panel 2 shows that a sizeable proportion of observations, roughly 25%, are cases in which an individual faces a completely new set of peers.

Across meetings, we also see variation in who the best performer is. An individual is the best performer in 6% of the total meetings attended. In 99% of the observations, an individual attends a group with a new top performer. More generally, there is variation in their relative ranks with respect to weight loss from one meeting to the next. Panel 3 provides a table that tabulates the transition matrix for *future* relative weight performance rank (as indicated by the columns) with *current* relative weight performance rank (as indicated by the rows) in meetings with 10 or more participants; akin to a heat map, we use darker shading to indicate larger values. As an example about how to interpret this figure, the top right element should be interpreted as a 5% probability that a top performer is the a 20th ranked performer in his or her next visit. This figure confirms that while some best performers continue to perform well in subsequent meetings, there are a large number of observations in which relative weight performance rank changes from one meeting to the next.

2.2.4 Instruments Used for Analysis

Finally, our empirical analysis makes use of instrumental variables to address the identification issues that we outline in the next section. Table 2 summarizes the main instruments we use for our analysis. Note that the distance variable is constructed using the physical distance between the meeting location the participant attended during the last period and participant's address. Furthermore, localized daily weather data is at the longitude-latitude level, and obtained from the National Centers for Environmental Information (NOAA). These summary statistics confirm that the distance and weather conditions faced by average, bottom, and top performers are indeed different.

3 Empirical Analysis of Heterogeneous Peer Effects

3.1 Main Specification

Our empirical strategy is an extension of the linear-in-means specification for peer effects (Brock and Durlauf, 2001; Manski, 1993). For a general overview of peer effects research in marketing, we refer readers to Hartmann et. al. (2008). We extend this specification by

Table 2: Summary Statistics for Distance and Daily Weather Instruments

Instrument	Mean	Std. Dev.
Distance to meeting (km)		
Average performer	22.1196	73.3447
Worst performer	24.3190	182.9587
Best performer	22.3350	169.1483
Weather instruments		
Precipitation (mm)	25.2323	70.0818
Max temperature (Fahrenheit)	185.368	101.9357
Min temperature (Fahrenheit)	84.4734	81.5069
Distance to meeting x *Min temperature*		
Average performer	2117.863	9000.345
Worst performer	2355.863	22864.62
Best performer	2129.498	21081.1
Distance to meeting x *Precipitation*		
Average performer	554.7244	6013.747
Worst performer	597.5546	12158.52
Best performer	556.8712	11542.3
Distance to meeting x *Max temperature*		
Average performer	4379.324	15697.87
Worst performer	4852.011	40708.42
Best performer	4416.226	37380.23

Note: Daily weather data is at the longitude-latitude level, and obtained from http://www.ncdc.noaa.gov/cdo-web.

incorporating more detailed information about the distribution of peer outcomes (i.e., best and worst performer outcomes), not just the average peer's performance, which then allow for these effects to be heterogeneous. The main specification we use is described as:

$$y_{it} = \alpha y_{it-1} + \gamma_1(y_{it-1}^{Avg} - y_{it-1}) + \gamma_2(y_{it-1}^{Worst} - y_{it-1}) + \gamma_3(y_{it-1}^{Best} - y_{it-1}) + \beta X_{it} + \mu_i + \mu_l + \mu_m + \varepsilon_{it}. \tag{1}$$

Here, y_{it} is the weight loss per day from meeting $t-1$ to t for individual i. Potential inertial effects are captured by α. Furthermore, peer effects are captured by $(\gamma_1, \gamma_2, \gamma_3)$, as we allow for the possibility that the difference between a peer's weight loss and i's past weight loss has an impact on i's current weight loss. The various peer outcomes are denoted by y_{it-1}^{Avg}, y_{it-1}^{Worst}, and y_{it-1}^{Best}, which capture the average, worst, and best weight loss among i's peers at a previous meeting $t-1$. The model also includes time-varying covariates of participant i in X_{it} to control for contextual effects, which contains the distance to goal, number of others at the previous meeting, number of days since the last meeting, the number of days since joining the weight loss program, as well as interactions between some of these variables and the peer weight loss outcomes. Instrumental variables used for our analysis are denoted by Z_{it}, which includes detailed information about each meeting participant's (i.e., individual, average performer, worst performer, best performer) physical distance to the meeting location, interacted with daily weather patterns (i.e., temperature, precipitation). Lastly, our model controls for any individual-level unobserved heterogeneity by including μ_i, location-level local unobserved heterogeneity by μ_l, and month fixed effects by μ_m. The month fixed effect can help control for national-level advertising and promotion campaigns.[3] We have provided a summary table with the list of key variables in Table 3 for the reader's convenience.

[3] Based on our conversations with company representatives in June 2014, the marketing activities were not targeted towards individuals, specific meetings, or locations. Therefore, a time dummy would be sufficient as a control for the marketing intensity.

Table 3: Summary Table With Variable Notation and Descriptions

Notation	Description
y_{it-1}	Last period weight loss per day (in kg)
y_{it-1}^{Avg}	Last period weight loss of the average performing peer (in kg)
y_{it-1}^{Worst}	Last period weight loss of the worst performing peer (in kg)
y_{it-1}^{Best}	Last period weight loss of the best performing peer (in kg)
X_{it}	Distance to goal, attendance count, experience, and relevant interactions
Z_{it}	Instruments including distance to meeting and local weather patterns
μ_i	User i fixed effect
μ_l	Location l fixed effect
μ_m	Month m fixed effect

Note: Distance variable is constructed using the physical distance between the meeting location the participant attended during the last period and participant's address.

Note that the main specification can be re-written as follows,

$$y_{it} = (\alpha - \gamma_1 - \gamma_2 - \gamma_3)y_{it-1} + \gamma_1 y_{it-1}^{Avg} + \gamma_2 y_{it-1}^{Worst} + \gamma_3 y_{it-1}^{Best} + \beta X_{it} + \mu_i + \mu_l + \mu_m + \varepsilon_{it},$$

which will ultimately be our estimation equation. By letting $\widetilde{\alpha} = \alpha - \gamma_1 - \gamma_2 - \gamma_3$, we can simplify the estimation equation further:

$$y_{it} = \widetilde{\alpha} y_{it-1} + \gamma_1 y_{it-1}^{Avg} + \gamma_2 y_{it-1}^{Worst} + \gamma_3 y_{it-1}^{Best} + \beta X_{it} + \mu_i + \mu_l + \mu_m + \varepsilon_{it}. \tag{2}$$

3.2 Identification

There are five main challenges to identification, namely selection into meetings, endogenous peer outcomes, individual unobserved heterogeneity, lagged dependent variables, and contextual confounds. This section will provide an outline of these issues, along with how we address such concerns in our empirical strategy.

3.2.1 Selection into Meetings

A primary identification concern is selection into meetings. As each participant can choose whether or not to attend a meeting, such decisions may be a function of their past weight loss successes or failures, among other factors. Therefore, changes in weight from one meeting to the next may be an artifact of individuals attending based on how much weight they had gained or lost in the past. For example, those who have experienced past successes may have a greater incentive to attend so as to gloat about their progress. We address these concerns by taking a few steps. First, we make sure to use the normalized measure of weight loss (i.e., weight loss per day). This way, weight loss measures are scaled and thus comparable across participants; at the very least, normalization helps us eliminate the mechanical relationship between days that separate meetings and weight loss. Second, we consider a conceptual robustness test to rule out the confounding nature of selection by analyzing the sensitivity of our results to sub-samples of participants based on the frequency to which they attend meetings (i.e., every month, every two weeks, every week). If our results were largely driven by decisions to attend meetings based on how much weight they had gained or lost in the past, then these results should disappear if we focus on participants who attend at regular frequencies (i.e., every month or week). Finally, based on the intuition from our sub-sample analysis, we consider a more direct way to address selection biases by including Heckman's (1979) selectivity correction term as a control function in the second-stage regression. We construct the selectivity correction term via Olsen (1980) by estimating a series of flexible linear probability models where the dependent variable is captured by a dummy that indicates whether a participant attended a meeting within a month, two weeks, or one week of the past meeting conditional on the weather and distance variables for the individual. An important advantage of using this approach (as opposed to probit) is that the normality assumption is not needed in Olsen's (1980) derivation of the estimator.

3.2.2 Endogenous Peer Outcomes

Selection into meetings will not only affect the individuals, but also the distribution of weight loss outcomes among a group of participants in a meeting. If people attend meetings to primarily show-off their past successes, then the group's weight loss success will likely be skewed to the right, while if people attend meetings to get back on a good weight loss trajectory (after suffering some past failures), then the distribution of weight loss among peers may be skewed the opposite direction. For this reason, we need an instrumental variable that could potentially shift attendance (and thus distribution of peer weight loss outcomes) in an exogenous manner. To construct these instruments, we make use of information about each individual's physical distance to the last meeting they attended from their residence along with localized weather conditions around their residence. These interacted variables then offer a plausible driver for variation in meeting attendance (and thus distribution of weight loss outcomes) among peers that is plausibly generated by exogenous factors that exclude the very selection concerns that affect both individuals and peers alike. The nice feature about these instruments is that they impact each member differently, as participants are exposed to different distance and weather conditions; this feature is especially important as we are interested in not only the average performer's impact, but also the best and worst performers.

3.2.3 Individual Unobserved Heterogeneity

Next, there is an issue regarding individual heterogeneity. Heterogeneity is a relevant as each participant may be inherently better or worse at losing weight due to psychological or physiological reasons. To control for individual-specific features, we include individual fixed effects in the estimation, akin to recent work about peer effects in marketing (e.g., Nair, Manchanda, and Bhatia, 2010). This approach is feasible as our data is rich in both the cross-sectional (i.e., number of participants) and time (i.e., number of repeat observations for the same individual over time) dimensions. Furthermore, the use of individual fixed effects relies to some extent variation in the composition of participants across meetings, which our empirical setting exhibits (see Section 2). Finally, heterogeneity may be related to an inherent tendency to attend meetings, so to some extent these fixed effects will also control for time-invariant selection into meetings.

3.2.4 Lagged Dependent Variables

An important feature to incorporate in our analysis is past weight loss (i.e., lagged dependent variable), as health habits are potentially inertial in nature. However, incorporating past weight loss will make estimation a little more complex, as the inclusion of a lagged dependent variable implies that past weight loss will be a function of unobserved factors. For example, some weight loss participants are inherently better at losing weight than others, and thus, our inferences about the inertial effects might be overestimated. For this reason, we rely on the Arellano and Bond (1991) method that makes use of lagged differences in the dependent variable as instruments for the most recent lagged term, as differencing between current and past dependent variables alone only eliminates permanent individual heterogeneity but not time varying individual shocks; for example, some individuals may perform better at losing weight in winter than others if they enjoy winter sports more than summer sports. Given the size of our data, we are able to observe each individual over time for a large number of observations, and thus, are able to make use of as many as 6 lagged differences when we explore robustness of this approach.

3.2.5 Contextual Confounds

One final source of bias may come from contextual effects. That is, there may be a common factor that all participants at a given meeting are exposed to. Some examples of these contextual factors may be if certain weight loss meeting locations provide better information (i.e., pamphlets they hand out at the beginning of session), facilities (i.e., meetings take place in room with better décor), employees (i.e., more experienced meeting leaders), or services (i.e., greater range of prepared food products available to participants). We accommodate for potential contextual effects by including location-specific fixed effects to control for these common factors that can affect every participant in a meeting.

3.3 Main Results

Table 4 provides the estimates from our first-stage estimation. The first four columns provide results from the first-stage linear regression of endogenous peer weight outcomes with instruments, while the latter three columns provide the results from the first-stage control function estimates for the selectivity correction term.

The first-stage results confirm that these instruments help explain some variation in the endogenous variables. For example, we see that distance has an impact on the average performers and worst performer's past weight loss. Also, rain has an impact on all of the endogenous variables. Furthermore, the minimum temperature affects the average performer and best performer's past weight loss, while the maximum temperature affects the average performer and worst performer's weight loss. Finally, we can see that the interactions between weather and distance metrics have an impact on these endogenous peer outcomes. We also note that weak instruments are unlikely to be an issue, as the F-statistics and R^2 markedly increase with the introduction of instruments (see Table 5). Note that in our table, we consider both univariate and multivariate F-tests, as there are multiple endogenous variables that need to be instrumented; for the multivariate F-tests, we use the approach by Sanderson and Windmeijer (2016).[4]

We also provide results from the estimates of our control function that is used to construct Heckman's (1979) selectivity correction term (see columns 5 to 7 in Table 4) via Olsen's (1980) method. The first column uses the correction term obtained from estimating the likelihood that a participant attends the current meeting within a month of the past meeting. The second column uses the correction term obtained from estimating the likelihood that

[4]We refer the reader to Leeflang, Wieringa, Bijmolt, and Pauwels (2015) for more details about the suggested tests of instrumental variable strength.

Table 4: First-Stage Linear Regression Results

	First-stage IV				First-stage control function		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Distance	-0.0000392***	0.0000230*	0.00000368	-0.00000211***	-0.00000942***	-0.0000164***	-0.0000398***
	(0.00000614)	(0.00000912)	(0.0000122)	(0.000000523)	(0.000000962)	(0.00000147)	(0.00000204)
Precipitation	0.0000311***	0.000152***	0.0000268**	-0.00000105*	-0.0000141***	-0.0000209***	-0.0000241***
	(0.00000212)	(0.00000769)	(0.00000934)	(0.000000425)	(0.000000781)	(0.00000119)	(0.00000166)
Temp max	0.000116***	0.000309***	0.0000148	0.0000210***	0.000181***	0.000258***	0.000244***
	(0.00000374)	(0.0000136)	(0.0000165)	(0.000000785)	(0.00000144)	(0.00000220)	(0.00000306)
Temp min	-0.000236***	0.0000134	-0.000518***	-0.00000864***	0.000178***	0.000125***	0.0000380***
	(0.00000474)	(0.0000171)	(0.0000208)	(0.000000989)	(0.00000182)	(0.00000278)	(0.00000386)
Distance x Temp min	0.000000282***	0.00000126***	-0.000000415***	8.38e-09	4.47e-08***	9.34e-08***	0.000000140***
	(6.33e-08)	(8.68e-08)	(0.000000117)	(5.41e-09)	(9.95e-09)	(1.52e-08)	(2.11e-08)
Distance x Precipitation	2.51e-08	7.01e-08	-5.29e-08	4.69e-10	-1.06e-08*	9.18e-09	1.49e-09
	(2.65e-08)	(4.73e-08)	(6.06e-08)	(2.43e-09)	(4.47e-09)	(6.83e-09)	(9.49e-09)
Distance x Temp max	-0.000000129*	-0.000000957***	0.000000398***	-1.80e-08***	-1.86e-08*	-7.33e-08***	-0.000000110***
	(5.26e-08)	(7.36e-08)	(9.99e-08)	(4.28e-09)	(7.87e-09)	(1.20e-08)	(1.67e-08)
Constant	0.213***	-1.486***	1.858***	0.0572***	0.990***	0.911***	0.729***
	(0.000597)	(0.00219)	(0.00266)	(0.000139)	(0.000255)	(0.000390)	(0.000541)
Observations	14395615	14395615	14395615	14395615	12397426	12397426	12397426

Note: Columns 1 to 4 provide results from the first-stage linear regression of endogenous peer weight outcomes with the instruments. The first column provides the results from first-stage linear regression of average performer's past weight loss on the weather-distance instruments. Analogously, the second, third and fourth columns provide results from first-stage linear regressions of worst performer's past weight loss, best performer's past weight loss, and individual's past weight loss per day on the weather-distance instruments respectively. Columns 5 to 7 provide the results from the first-stage control function estimates for the selectivity correction term. The fifth column uses the correction term obtained from estimating the likelihood participant attended current meeting within a month of the past meeting. The sixth column uses the correction term obtained from estimating the likelihood participant attended current meeting within 2 weeks of the last meeting. The seventh column uses the correction term obtained from estimating the likelihood participant attended current meeting within 1 week of the last meeting.

Table 5: Strength of the Instrumental Variables

	(1)	(2)	(3)	(4)
Univariate F-statistics				
Exogenous variables with instruments	69264	32823	9861	183634
Exogenous variables	51386	18700	4243	99999
Multivariate F-statistics				
Exogenous variables with instruments	3820	24744	18966	31873
Exogenous variables	1679	1442	1513	1964
R^2				
Exogenous variables with instruments	0.0379	0.0137	0.0104	0.1786
Exogenous variables	0.0097	0.0035	0.0008	0.0436

Note: The first column provides the F-statistics and R^2 from the first-stage regressions of average performer's past weight loss. Analogously, the second, third and fourth columns provide the corresponding results from regressions of worst performer's past weight loss, best performer's past weight loss, and individual's past weight loss per day. For the multivariate F-statistics, we report the weak instrument F-test statistic by Sanderson and Windmeijer (2016), as there are multiple endogenous variables that are instrumented.

a participant attends the current meeting within 2 weeks of the last meeting. The third column uses the correction term obtained from estimating the likelihood that a participant attends a current meeting within 1 week of the last meeting. These results confirm that the weather and distance metrics help explain some of the variation in the likelihood of frequent meeting attendance.

The key findings about the effect of peers on weight loss progress are found in Table 6. The Online Appendix provides additional robustness checks, including the conceptual test we use to rule out biases from selection. Columns 1 to 3 correspond to the correction terms obtained from estimating the likelihood participant attended current meeting within a month of the past meeting, within two weeks of the last meeting and within one week of the last meeting, respectively. For all of the specifications, we make use of the lagged instruments as suggested by Arellano and Bond (1991). Our results are invariant to how many lags are used, which provides further support that the estimates and standard errors are unlikely to be biased by any serially correlation that may exist (see Online Appendix). Since we have interaction effects in the specifications, we make use of mean-centered continuous variables. Mean-centering affects the interpretation of the results, in that the intercepts are more meaningful now - the inferred moderating effects reflect differences in weight loss outcomes for an individual facing average levels of individual and peer performance. Before summarizing the peer effects, we would like to point out that there are some inertial effects in weight loss. That is, an individual's past weight loss is associated with an increase in subsequent weight loss. This finding suggests potential long-run implications of interventions that affect current weight loss outcomes. Focusing the peer effects, we see in the baseline specification that weight loss for the average peer leads to individual weight gain, as a 1 kg increase in the average peer performer's weight loss is associated with an individual's decrease in weight loss by about 0.02 kg. In contrast, we see that weight loss by the top performer leads to increased individual weight loss, as a 1 kg increase in the top performer's weight loss is associated with an individual's increase in weight loss by about 0.01 kg. While

Table 6: Second-Stage Linear Regression Results

	(1)	(2)	(3)
Last period weight loss per day	0.0505***	0.0496***	0.0490***
	(0.000721)	(0.000736)	(0.000751)
Distance to goal	0.0272***	0.0272***	0.0272***
	(0.0000402)	(0.0000403)	(0.0000403)
Distance to goal x Weight loss per day	0.00447***	0.00447***	0.00447***
	(0.00000214)	(0.00000214)	(0.00000215)
Average performer	-0.0240***	-0.0234***	-0.0228***
	(0.000702)	(0.000710)	(0.000722)
Worst performer	-0.00736***	-0.00679***	-0.00664***
	(0.000499)	(0.000493)	(0.000484)
Best performer	0.00928***	0.00877***	0.00862***
	(0.000449)	(0.000444)	(0.000437)
Attendance count	0.000104***	0.000105***	0.000105***
	(0.00000473)	(0.00000473)	(0.00000473)
Average performer x Attendance count	0.000436***	0.000439***	0.000442***
	(0.0000283)	(0.0000283)	(0.0000283)
Worst performer x Attendance count	-0.0000834***	-0.0000835***	-0.0000837***
	(0.00000606)	(0.00000605)	(0.00000605)
Best performer x Attendance count	-0.0000299***	-0.0000297***	-0.0000297***
	(0.00000450)	(0.00000450)	(0.00000449)
Experience	0.0118***	0.0117***	0.0117***
	(0.000272)	(0.000272)	(0.000272)
Average performer x Experience	0.000373***	0.000371***	0.000370***
	(0.0000445)	(0.0000445)	(0.0000444)
Worst performer x Experience	0.000192***	0.000192***	0.000191***
	(0.0000246)	(0.0000245)	(0.0000245)
Best performer x Experience	-0.000158***	-0.000157***	-0.000157***
	(0.0000225)	(0.0000225)	(0.0000225)
Selection correction	-0.00729***	-0.0109***	-0.0144***
	(0.00215)	(0.00181)	(0.00197)
Constant	-0.188***	-0.185***	-0.184***
	(0.00254)	(0.00215)	(0.00198)
Observations	12376990	12376990	12376990

Standard errors in parentheses

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Note: The first column uses the correction term obtained from estimating the likelihood participant attended current meeting within a month of the past meeting. The second column uses the correction term obtained from estimating the likelihood participant attended current meeting within 2 weeks of the last meeting. The third column uses the correction term obtained from estimating the likelihood participant attended current meeting within 1 week of the last meeting.

the magnitudes of these effects seem small, we note that a large number of observations (about 41%) involve participants gaining weight; therefore, any positive impact on weight loss success would be valuable to participants.

3.3.1 Discussion

As we focus on heterogeneity in peer performance outcomes, we not only complement the past literature in marketing that has focused on other dimensions of heterogeneity (i.e., spatial proximity, observable physical characteristics, intra-group relationships, opinion leadership, network tie strength), but are also able to contribute towards the literature about social comparison theory. In particular, the discrepancy between best and average performer effects could potentially be explained by an adaptive function related to *upward* comparisons. The negative average performer effects would then be consistent with the idea that peers who perform well may cause individuals to feel inadequate and that they themselves cannot achieve such levels of performance (e.g., Rogers and Feller, 2016). Consequently, individuals may attempt to respond in a defensive manner whenever someone else outperforms them. One strategy that has been alluded to in the social comparison literature is to self-handicap and choose an obviously superior peer as a comparison target (e.g., Shepperd and Taylor, 1999). It then seems plausible that this self-handicapping is less effective when the comparison target is an average performer, as opposed to an obviously superior top performer. By self-handicapping, individuals can exploit upward drives in their comparisons with top performers, while at the same time, counteract their feelings of under-performance (relative to peers). Our findings suggest that top performing peers provide better motivation to weight loss participants than average performing peers, which appears consistent with the theoretical predictions about upward comparisons and self-handicapping.

In summary, the evidence of heterogeneous peer effects has direct implications on meeting design for the commercial weight loss program. There are two main dimensions of our study's managerial implications. The first dimension of meeting design that our results may impact is *content*. For example, the meeting leaders can use the weight loss successes of best performers provide inspiration to the group, and perhaps avoid using the average performer's success as the benchmark. By focusing attention on the top performer, the weight loss program can better motivate participants towards sustainable weight loss. The second dimension of meeting design that may be impacted by our findings is *composition*. The weight loss program can form groups of meeting participants that would maximize the encouraging effects of best performers and minimize the discouraging effects of average performers. By improving the perceived performance of the weight loss program via these meeting design strategies, the firm can maintain a high level of customer satisfaction and engagement.

Based on our results, we now discuss possible moderating effects, which may allow us to refine the managerial implications. These interactions will help the weight loss program determine under which contexts the peers are more or less encouraging/discouraging. The first interaction we look at is between the peer effects and attendance count. The average performer effect appears to be slightly less negative as the group size increases, while the interaction between group size and the best performer effect is small in magnitude and statistically significant. An implication of this result is that when the group size is large, announcing the successes of the average performer appears to be less detrimental towards an individual's future performance. Furthermore, the second interaction investigates whether the heterogeneous peer effects vary with an individual's experience in the weight loss program. Measuring experience as the number of days the individual has been a member of the weight loss program, we show that the average, worst, and best performer effects do not vary much with an individual's own experience, though the best performer effect is slightly smaller for individuals who are more experienced. We also tried defining experience based on the total number of meetings the participant has attended thus far, and the results are qualitatively the same. The moderating effects for experience reveal that highlighting the successes of best performers may be slightly more effective if the group consists primarily of inexperienced attendees.

While we focus primarily on the heterogeneous peer effects, future work could help refine further the possible mechanisms behind the inferred patterns associated with best performer peer effects. Based on the framework by Van den Bulte and Lilien (2001), one can break-down various causal mechanisms behind social influence. The first possible mechanism is information transfer, such as social learning or vicarious learning. We argue that information transfer is unlikely to be a driver of the social effects as the top performer effects do not vary much with an individual's own experience. The second possible mechanism is performance network effect. Based on how weight loss programs work, it seems unlikely that one peer's weight loss will lead to a direct physical benefit to those attending the same meeting. The third possible mechanism is normative pressure. We believe this mechanism is an unlikely driver behind the peer effects as the meeting attendees are acquaintances with one another, and thus, approval among peers may not be as valuable than in cases in which members of a group are actually friends; though, a limitation of this argument is that descriptive norms may have an impact even if members in the group are not friends with one another (e.g., Cialdini, Reno, and Kallgren, 1990), such that norms are defined by what is typical or normal (i.e., what most people do). Finally, the remaining mechanism is competitive concern. We believe this mechanism may potentially fit with our empirical findings, in that the discouraging average peer effect is dampened by the number of peers attending the meeting. In summary, the main effects as well as their interactions with various proxies suggest that social comparisons play a role behind the peer effects. Given the resemblance between social comparisons and competitive concerns, we speculate that competitive concerns would be a mechanism that fits best with our empirical findings, though future work is needed to confirm that this is indeed the case.

4 Conclusion

Our study investigates the role of social interactions in inducing healthier behavior. We infer heterogeneous peer effects using data from a large commercial weight loss program. The findings indicate that while the weight loss among average peers does not lead to individual weight loss, weight loss among best performing peers has a positive impact on weight loss progress. In summary, our research suggests opportunities to improve the design of meetings by highlighting the best performer's successes (i.e., meeting *content*), or by forming groups that minimize the discouraging average performer's effect while maximizing the encouraging best performer's effect (i.e., meeting *composition*). Although this is beyond the scope of our study, we see potential in future work to explore analytically what the optimal composition would be.

From a health management perspective, future research could also investigate the interaction between peer effects and urgency of weight loss. As Ma, Ailawadi, and Grewal (2013) demonstrate that consumption of healthy foods increase in response to diabetes diagnosis, the adoption of healthful behavior appears to be affected by medical conditions. It would be interesting to see if the encouraging top performer effects are particularly helpful at motivating those who are in greatest need of losing weight in a short amount of time. Since weight maintenance is a key preventative measure against diabetes, such a finding would have health implications above and beyond the specific weight loss context we study.

Finally, our research insights can be applied to scenarios well beyond weight management. For example, some contexts in which identifying the ideal role model may be important include sustainable technology diffusion, sales-force motivation, as well as education design. In the adoption of solar panel technologies (e.g., Bollinger and Gillingham, 2012; Kraft-Todd et. al., 2017), one could explore the role of leaders in a community to foster adoption across households. To augment the sales-force control systems (Cravens et. al., 1993), our findings may suggest which employees a company should highlight in order to motivate others. For this reason, we see our paper's insights as being complementary with Chan, Li, and Pierce (2014b) and Mas and Moretti (2009), as they provide evidence of peer effects in sales productivity. Similarly, in a classroom setting (Imberman, Kugler, and Sacerdote, 2012), teachers can highlight the successes of top performers, rather than the class as a whole in order to motivate students. Furthermore, the increasing role of gamification and social interactions in many mobile applications (Hofacker et. al., 2016) may yield fruitful settings to study who and how peer successes should be highlighted.

References

[1] Anderson, E., Fornell, C., and Lehmann, D. 1994. "Customer Satisfaction, Market Share, and Profitability: Findings From Sweden." *Journal of Marketing* 58: 53-66.

[2] Aral, S., and Walker, D. 2014. "Tie Strength, Embeddedness, and Social Influence: A Large-Scale Networked Experiment." *Management Science* 60: 1352-1370.

[3] Arellano, M., and Bond, S. 1991. "Some tests of specification for panel data: Montel Carlo evidence and an application to employment equations." *Review of Economic Studies* 58: 277-297.

[4] Bell, D., and Song, S. 2007. "Neighborhood effects and trial on the Internet: Evidence from online grocery retailing." *Quantitative Marketing and Economics* 5: 361-400.

[5] Blanton, H., Buunk, B., Gibbons, F., and Kuyper, H. 1999. "When Better-Than-Others Compare Upward: Choice of Comparison and Comparative Evaluation as Independent Predictors of Academic Performance." *Journal of Personality and Social Psychology* 76: 420-430.

[6] Bollinger, B., and Gillingham, K. 2012. "Peer Effects in the Diffusion of Solar Photovoltaic Panels." *Marketing Science* 31: 900-912.

[7] Brewer, M., and Weber, J. 1994. "Self-Evaluation Effects of Interpersonal Versus Intergroup Social Comparison." *Journal of Personality and Social Psychology* 66: 268-275.

[8] Brock, W., and Durlaf, S. 2001. "Interaction-based models." *Handbook of Econometrics* 5: 3297-3380.

[9] Buunk, A., and Gibbons, F. 2007. "Social comparison: The end of a theory and the emergence of a field." *Organizational Behavior and Human Decision Processes* 102: 3-21.

[10] Chan, T., Li, J., and Pierce, L. 2014b. "Learning from Peers: Knowledge Transfer and Sales Force Productivity Growth." *Marketing Science* 33: 463-484.

[11] Choi, J., Hui, S., and Bell, D. 2010. "Spatiotemporal Analysis of Imitation Behavior Across New Buyers at an Online Grocery Retailer." *Journal of Marketing Research* 42: 75-89.

[12] Christakis, N.A., and Fowler, J.H. 2008. "The spread of obesity in a large social network over 32 years." *New England Journal of Medicine* 358: 2249-2258.

[13] Cialdini, R., Reno, R., and Kallgren, C. 1990. "A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places." *Journal of Personality and Social Psychology* 58: 1015-1026.

[14] Cravens, D., Ingram, T., LaForge, R., and Young, C. 1993. "Behavior-Based and Outcome-Based Salesforce Control Systems." *Journal of Marketing* 57: 47-59.

[15] Fitzsimons, G., and Fishbach, A. 2010. "Shifting Closeness: Interpersonal Effects on Personal Goal Progress." *Journal of Personality and Social Psychology* 98: 535-549.

[16] Gardete, P. 2014. "Understanding Social Effects in the In-Flight Marketplace: Characterization and Managerial Implications." *Journal of Marketing Research*, forthcoming.

[17] Godes, D., and Mayzlin, D. 2009. "Firm-Created Word-of-Mouth Communication: Evidence from a Field Test." *Marketing Science* 28: 721-739.

[18] Hagen, L., Krishna, A., and McFerran, B. 2016. "Rejecting Responsibility: Low Physical Involvement in Obtaining Food Promotes Unhealthy Eating." *Journal of Marketing Research*, forthcoming.

[19] Hartmann et. al. 2008. "Modeling social interactions: Identification, empirical methods and policy implications." *Marketing Letters* 19: 287-304.

[20] Heckman, J. 1979. "Sample selection bias as a specification error." *Econometrica* 47: 153-161.

[21] Hofacker, C., de Ruyter, K., Lurie, N., Manchanda, P., and Donaldson, J. 2016. "Gamification and Mobile Marketing Effectiveness." *Journal of Interactive Marketing* 34: 25-36.

[22] Imberman, S., Kugler, A., and Sacerdote, B. 2012. "Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees." *American Economic Review* 102: 2048-2082.

[23] Iyengar, R., Van den Bulte, C., and Valente, T. 2011. "Opinion Leadership and Social Contagion in New Product Diffusion." *Marketing Science* 30: 195-212.

[24] Karlsson, N., Loewenstein, G., and Seppi, D. 2009. "The ostrich effect: Selective attention to information." *Journal of Risk and Uncertainty* 38: 95-115.

[25] Kraft-Todd, G., Bollinger, B., Gillingham, K., Lamp, S., and Rand, D. 2017. "Credibility-Enhancing Displays Promote the Provision of a Non-Normative Public Good." Working paper, Yale University.

[26] Kremer, M., and Levy, D. 2008. "Peer Effects and Alcohol Use among College Students." *Journal of Economic Perspectives* 22: 189-206.

[27] Kumar, V., Umashankar, N., Kim, K., and Bhagwat, Y. 2014. "Assessing the Influence of Economic and Customer Experience Factors on Service Purchase Behaviors." *Marketing Science* 33: 673-692.

[28] Leeflang, P., Wieringa, J., Bijmolt, T., and Pauwels, K. 2015. *Modeling Markets: Analyzing Marketing Phenomena and Improving Marketing Decision Making*. Springer: New York.

[29] Lockwood, P., Jordan, C., and Kunda, Z. 2002. "Motivation by Positive or Negative Role Models: Regulatory Focus Determines Who Will Best Inspire Us." *Journal of Personality and Social Psychology* 83: 854-864.

[30] Ma, Y., Ailawadi, K., and Grewal, D. 2013. "Soda Versus Cereal and Sugar Versus Fat: Drivers of Healthful Food Intake and the Impact of Diabetes Diagnosis." *Journal of Marketing* 77: 101-120.

[31] Manchanda, P., Xie, Y., and Youn, N. 2008. "The Role of Targeted Communication and Contagion in Product Adoption." *Marketing Science* 27: 961-976.

[32] Manski, C. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic Studies* 60: 531-542.

[33] Mas, A., and Moretti, E. 2009. "Peers at Work." American Economic Review 99: 112-145.

[34] McFerran, B., Dahl, D., Fitzsimons, G., and Morales, A. 2010a. "I'll Have What She's Having: Effects of Social Influence and Body Type on the Food Choices of Others." *Journal of Consumer Research* 36: 915-929.

[35] McFerran, B., Dahl, D., Fitzsimons, G., and Morales, A. 2010b. "Might an overweight waitress make you eat more? How the body type of others is sufficient to alter our food consumption." *Journal of Consumer Psychology* 20: 146-151.

[36] Morse, S., and Gergen, K. 1970. "Social Comparison, Self-Consistency, and the Concept of Self." *Journal of Personality and Social Psychology* 16: 148-156.

[37] Nair, H., Manchanda, P., and Bhatia, T. 2010. "Asymmetric Social Interactions in Physician Prescription Behavior: The Role of Opinion Leaders." *Journal of Marketing Research* 47: 883-895.

[38] Narayanan, S., and Nair, H. 2012. "Estimating causal installed-base effects." *Journal of Marketing Research* 50: 70-94.

[39] Narayan, V., Rao, V., and Saunders, C. 2011. "*How Peer Influence Affects Attribute Preferences: A Bayesian Updating Mechanism.*" Marketing Science 30: 368-384.

[40] Olsen, R. 1980. "A Least Squares Correction for Selectivity Bias." *Econometrica* 48: 1815-1820.

[41] Park, H-M., and Manchanda, P. 2014. "When Harry Bet with Sally: An Empirical Analysis of Multiple Peer Effects in Casino Gambling Behavior." *Marketing Science*, forthcoming.

[42] Pelham, B., and Wachsmuth, J. 1995. "The Waxing and Waning of the Social Self: Assimilation and Contrast in Social Comparison." *Journal of Personality and Social Psychology* 5, 825-838.

[43] Rogers, T., and Feller, A. 2016. "Discouraged by Peer Excellence: Exposure to Exemplary Peer Performance Causes Quitting." *Psychological Science*, 27: 365-374.

[44] Sanderson, E., and Windmeijer, F. 2016. "A Weak Instrument F-Test in Linear IV Models with Multiple Endogenous Variables." *Journal of Econometrics* 190: 212-221.

[45] Shepperd, J., and Taylor, K. 1999. "Ascribing advantages to social comparison targets." *Basic and Applied Social Psychology* 21: 103-117.

[46] Van den Bulte, C., and Lilien, G. 2001. "Medical Innovation Revisited: Social Contagion versus Marketing Effort." *American Journal of Sociology* 106: 1409-1435.

[47] Yang, S., Hu, M., Winer, R., and Assael, H. 2012. "An Empirical Study of Word-of-Mouth Generation and Consumption." *Marketing Science* 31: 952-963.

[48] Yang, S., Narayan, V., and Assael, H. 2006. "Estimating the Interdependence of Television Program Viewership Between Spouses: A Bayesian Simultaneous Equation Model." *Marketing Science* 25: 336-349.

Online Appendix (Not for Publication)

Conceptual Test to Establish Robustness from Selection

To further assuage our concerns about selection biases, we consider a conceptual robustness check, in additional to using the scaled dependent variable and instruments for the individual's past weight loss per day. For the conceptual robustness check, we consider scenarios in which weight loss participants are plausibly selecting into meetings. The scenarios we consider are those in which participants attend the meetings on a regular basis. Based on this scenario definition, we consider a few sub-samples of varying levels of selection. In particular, the sub-samples we use include individuals who have regularly attended meetings every month, every two weeks, and every week. The results from our analysis of these sub-samples are displayed in columns (2) to (4) of Tables 7 and 8 respectively. Column 1 shows the results from the full-sample for comparison purposes. Note that the comparison across these tables alone is informative as it may suggest that the assumption about uncorrelated errors over time is well justified. Our robustness checks from the sub-sample analysis confirms that the negative average performer effect and positive best performer effect remain unchanged qualitatively across all sub-samples with varying degrees of selection.

Multicollinearity

In this section, we explore robustness of our results to multicollinearity. If the various peer effect metrics are strongly correlated with one another, then there may be implications about the design of meetings (i.e., how information is shared during meetings), as some of our meeting design suggestions require being able to infer one type of peer effect while holding the others fixed. Our analysis in this section will reveal that such issues are unlikely to have an impact on our final conclusions. Table 9 provides the baseline specification, as we include each of the peer effect metrics and attendance count in a sequential manner. The table

Table 7: Second-Stage Regression with Instrumental Variables

	(1)	(2)	(3)	(4)
Last period weight loss per day	0.0506***	0.0472***	0.0329***	0.0131***
	(0.000720)	(0.000724)	(0.000742)	(0.00123)
Distance to goal	0.0272***	0.0267***	0.0247***	0.0147***
	(0.0000402)	(0.0000404)	(0.0000423)	(0.0000685)
Distance to goal x Last period weight loss per day	0.00447***	0.00460***	0.00499***	0.00654***
	(0.00000214)	(0.00000219)	(0.00000244)	(0.00000565)
Average performer	-0.0245***	-0.0227***	-0.0155***	-0.00893***
	(0.000686)	(0.000722)	(0.000840)	(0.00180)
Worst performer	-0.00819***	-0.00640***	0.000272	-0.00271**
	(0.000435)	(0.000440)	(0.000480)	(0.000909)
Best performer	0.0100***	0.00803***	0.00199***	0.00402***
	(0.000394)	(0.000400)	(0.000441)	(0.000857)
Attendance count	0.000103***	0.000107***	0.0000928***	0.0000955***
	(0.00000473)	(0.00000481)	(0.00000534)	(0.0000103)
Average performer x Attendance count	0.000438***	0.000520***	0.000663***	0.000404***
	(0.0000283)	(0.0000294)	(0.0000338)	(0.0000699)
Worst performer x Attendance count	-0.0000837***	-0.0000653***	-0.0000589***	0.0000276*
	(0.00000606)	(0.00000619)	(0.00000698)	(0.0000135)
Best performer x Attendance count	-0.0000304***	-0.0000357***	-0.0000393***	-0.0000798***
	(0.00000450)	(0.00000465)	(0.00000553)	(0.0000116)
Experience	0.0120***	0.0116***	0.0103***	0.0101***
	(0.000268)	(0.000269)	(0.000286)	(0.000292)
Average performer x Experience	0.000373***	0.000371***	0.000349***	0.000643***
	(0.0000445)	(0.0000460)	(0.0000501)	(0.000122)
Worst performer x Experience	0.000192***	0.000174***	0.000142***	0.0000833
	(0.0000246)	(0.0000248)	(0.0000273)	(0.0000559)
Best performer x Experience	-0.000158***	-0.000141***	-0.000113***	-0.000125*
	(0.0000225)	(0.0000227)	(0.0000245)	(0.0000535)
Constant	-0.196***	-0.193***	-0.183***	-0.0841***
	(0.00107)	(0.00106)	(0.00109)	(0.000607)
Observations	12376990	12037102	10158025	2718078

Standard errors in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Note: The sub-samples we use include individuals who have regularly attended meetings every month, every two weeks, and every week. The results from our analysis of these sub-samples are displayed in columns (2) to (4). Column 1 shows the results from the full-sample for comparison purposes.

Table 8: Second-Stage Regression with Instrumental Variables and Location Fixed Effects

	(1)	(2)	(3)	(4)
Last period weight loss per day	0.0506***	0.0472***	0.0329***	0.0131***
	(0.000720)	(0.000724)	(0.000742)	(0.00123)
Distance to goal	0.0272***	0.0267***	0.0247***	0.0147***
	(0.0000402)	(0.0000404)	(0.0000423)	(0.0000685)
Distance to goal x Last period weight loss per day	0.00447***	0.00460***	0.00499***	0.00654***
	(0.00000214)	(0.00000219)	(0.00000244)	(0.00000565)
Average performer	-0.0245***	-0.0227***	-0.0155***	-0.00895***
	(0.000686)	(0.000722)	(0.000840)	(0.00180)
Worst performer	-0.00819***	-0.00640***	0.000272	-0.00269**
	(0.000435)	(0.000440)	(0.000480)	(0.000909)
Best performer	0.0100***	0.00803***	0.00199***	0.00401***
	(0.000394)	(0.000400)	(0.000441)	(0.000857)
Attendance count	0.000103***	0.000107***	0.0000927***	0.0000953***
	(0.00000473)	(0.00000481)	(0.00000534)	(0.0000103)
Average performer x Attendance count	0.000438***	0.000520***	0.000663***	0.000405***
	(0.0000283)	(0.0000294)	(0.0000338)	(0.0000699)
Worst performer x Attendance count	-0.0000837***	-0.0000653***	-0.0000590***	0.0000275*
	(0.00000606)	(0.00000619)	(0.00000698)	(0.0000135)
Best performer x Attendance count	-0.0000304***	-0.0000356***	-0.0000392***	-0.0000798***
	(0.00000450)	(0.00000465)	(0.00000553)	(0.0000116)
Experience	0.0120***	0.0116***	0.0103***	0.0101***
	(0.000268)	(0.000269)	(0.000286)	(0.000292)
Average performer x Experience	0.000373***	0.000371***	0.000349***	0.000644***
	(0.0000445)	(0.0000460)	(0.0000501)	(0.000122)
Worst performer x Experience	0.000192***	0.000174***	0.000142***	0.0000825
	(0.0000246)	(0.0000248)	(0.0000273)	(0.0000559)
Best performer x Experience	-0.000158***	-0.000141***	-0.000113***	-0.000125*
	(0.0000225)	(0.0000227)	(0.0000245)	(0.0000535)
Constant	-0.202***	-0.199***	-0.189***	-0.0902***
	(0.00206)	(0.00208)	(0.00229)	(0.00447)
Observations	12376990	12037102	10158025	2718078

Standard errors in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Note: The sub-samples we use include individuals who have regularly attended meetings every month, every two weeks, and every week. The results from our analysis of these sub-samples are displayed in columns (2) to (4). Column 1 shows the results from the full-sample for comparison purposes.

Table 9: The Inclusion of the Peer Variables and Attendance Count in Sequential Manner

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Average performer	-0.0110***	-0.00984***	-0.0167***				-0.0169***
	(0.000413)	(0.000478)	(0.000549)				(0.000549)
Worst performer		-0.000909***	-0.0101***	-0.00294***		-0.00744***	-0.00937***
		(0.000191)	(0.000401)	(0.000164)		(0.000391)	(0.000402)
Best performer			0.00931***		0.00143***	0.00395***	0.00866***
			(0.000357)		(0.000131)	(0.000311)	(0.000358)
Attendance count							0.000120***
							(0.00000427)
Constant	-0.145***	-0.145***	-0.144***	-0.145***	-0.145***	-0.144***	-0.145***
	(0.000309)	(0.000309)	(0.000311)	(0.000308)	(0.000308)	(0.000310)	(0.000311)
Observations	12376990	12376990	12376990	12376990	12376990	12376990	12376990

Standard errors in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Note: For simpler interpretation, these specifications do not include the interaction terms.

confirms that the sign and statistical significance for the average, worst, and best performer effects do not change as each peer metric is successively included into the specification. We note that even by using the most conservative estimates from these columns, the average performer, worst performer, and best performer effects have a non-negligible and statistically significant impact on an individual's subsequent weight loss.

Local Traffic Instruments

As further robustness, we consider specifications that make use of additional instrumental variables. In particular, we add local traffic conditions as instruments for our analysis. This data is obtained from past research by Duranton and Turner (2011) about road congestion in the United States. In particular, we make use of the averaged daily traffic for all MSA interstates, highways and roads. Table 10 provides the results from this analysis. The results confirm that most of our main findings about the heterogeneous peer effects hold, even when this new instrumental variable is added to the first-stage estimation. In particular, the average performer effect remains statistically significant up until we focus only on participants who attend meetings on a weekly basis. Furthermore, the top performer effect remains positive and statistically significant in all sub-samples. We note, however, that this specification

Table 10: Second-Stage Regression with Additional Instrumental Variables

	(1)	(2)	(3)	(4)
Last period weight loss per day	0.0502***	0.0468***	0.0324***	0.0129***
	(0.000719)	(0.000722)	(0.000740)	(0.00123)
Distance to goal	0.0272***	0.0267***	0.0247***	0.0147***
	(0.0000402)	(0.0000403)	(0.0000423)	(0.0000684)
Distance to goal x Last period weight loss per day	0.00447***	0.00460***	0.00499***	0.00654***
	(0.00000214)	(0.00000219)	(0.00000244)	(0.00000564)
Average performer	-0.0169***	-0.0145***	-0.00560***	-0.000317
	(0.000549)	(0.000571)	(0.000649)	(0.00134)
Worst performer	-0.00937***	-0.00728***	-0.000678	-0.00179*
	(0.000402)	(0.000407)	(0.000443)	(0.000835)
Best performer	0.00866***	0.00683***	0.00104**	0.00179*
	(0.000358)	(0.000362)	(0.000396)	(0.000761)
Attendance count	0.000120***	0.000117***	0.000103***	0.0000666***
	(0.00000427)	(0.00000433)	(0.00000475)	(0.00000916)
Constant	-0.145***	-0.145***	-0.141***	-0.0843***
	(0.000311)	(0.000314)	(0.000340)	(0.000598)
Observations	10390068	9783509	8094044	2361522

Standard errors in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Note: The sub-samples we use include individuals who have regularly attended meetings every month, every two weeks, and every week. The results from our analysis of these sub-samples are displayed in columns (2) to (4). Column 1 shows the results from the full-sample for comparison purposes.

has not been chosen as the baseline, as the local traffic data from Duranton and Turner (2011) does not have full coverage across all of the cities that have weight loss meetings.

Potential Serial Correlation

In this analysis, we explore robustness of our specifications to potential serial correlation. To demonstrate this robustness, we consider variants of our baseline specification by implementing the Arellano and Bond (1991) estimator with more lagged differences (i.e., $\triangle y_{it-2} = y_{it-2} - y_{it-3}$) as instruments. Increasing the number of lagged differences is one method to correct for potential serial correlation biases in the estimates and standard errors. Our results from this analysis can be found in Table 11. Columns 1 to 4 provide the results for specifications with 3 to 6 lagged differences respectively. The analysis confirms that even when a large number of lagged differences are included, the standard errors as well as estimates for the peer effects remain virtually the same as in the baseline. Given the large size of our data, we imagine that serial correlation in itself is unlikely to have a material impact on the inferred magnitudes and precision of our estimates.

References

[1] Duranton, G., and Turner, M. 2011. "The Fundamental Law of Road Congestion: Evidence from US Cities." *American Economic Review* 101: 2616-2652.

Table 11: Second-Stage Regression with Different Number of Lagged Differences

	(1)	(2)	(3)	(4)
Last period weight loss per day	0.0517***	0.0491***	0.0497***	0.0483***
	(0.000711)	(0.000716)	(0.000736)	(0.000747)
Distance to goal	0.0281***	0.0288***	0.0301***	0.0314***
	(0.0000434)	(0.0000466)	(0.0000504)	(0.0000543)
Last period weight loss per day x Distance to goal	0.00455***	0.00464***	0.00466***	0.00467***
	(0.00000224)	(0.00000235)	(0.00000245)	(0.00000254)
Average performer	-0.0246***	-0.0232***	-0.0246***	-0.0235***
	(0.000728)	(0.000784)	(0.000836)	(0.000888)
Worst performer	-0.00749***	-0.00686***	-0.00688***	-0.00725***
	(0.000457)	(0.000477)	(0.000502)	(0.000527)
Best performer	0.00974***	0.00899***	0.00909***	0.00957***
	(0.000415)	(0.000435)	(0.000462)	(0.000490)
Attendance count	0.000104***	0.000117***	0.000124***	0.000112***
	(0.00000503)	(0.00000530)	(0.00000563)	(0.00000598)
Average performer x Attendance count	0.000457***	0.000406***	0.000420***	0.000327***
	(0.0000304)	(0.0000327)	(0.0000351)	(0.0000375)
Worst performer x Attendance count	-0.0000940***	-0.0000735***	-0.0000735***	-0.0000778***
	(0.00000657)	(0.00000701)	(0.00000753)	(0.00000805)
Best performer x Attendance count	-0.0000305***	-0.0000344***	-0.0000328***	-0.0000285***
	(0.00000497)	(0.00000545)	(0.00000594)	(0.00000649)
Experience	0.0140***	0.0123***	0.00894***	0.00723***
	(0.000328)	(0.000526)	(0.000229)	(0.000421)
Average performer x Experience	0.000388***	0.000401***	0.000385***	0.000363***
	(0.0000462)	(0.0000478)	(0.0000509)	(0.0000528)
Worst performer x Experience	0.000192***	0.000186***	0.000201***	0.000223***
	(0.0000262)	(0.0000274)	(0.0000291)	(0.0000310)
Best performer x Experience	-0.000162***	-0.000163***	-0.000171***	-0.000183***
	(0.0000237)	(0.0000248)	(0.0000261)	(0.0000276)
Constant	-0.211***	-0.209***	-0.204***	-0.175***
	(0.00134)	(0.00222)	(0.000934)	(0.000425)
Observations	11348884	10419683	9573622	8798315

Standard errors in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Note: Columns 1 to 4 provide the results with 3 to 6 lagged differences respectively.