The Effect of Incentives on Physical Activity: an Alaska Experiment

Mouhcine Guettabi*

November 1, 2017

Abstract

In the Fall of 2016, we randomized 13 schools into treatment and control schools.We asked teachers in both groups to allow students to use the SQORD normally without issuing any challenges or competitions during the month of September. Starting October 14th, students in both groups were told by their teachers that they will be setting daily goals at 70,000 daily SQORD points for both boys and girls. These numbers roughly approximate the points that translate to 60 minutes of moderate-vigorous physical activity in a day. They were also told that on a weekly basis, the target would be considered as achieved if they reached the mark on at least 3 out of 6 days. In addition to these instructions, students in the treatment group were promised rewards if they reach the target. I find that incentives resulted in a higher share of students achieving their goals. This effect, however, was not consistent across schools.

Keywords: Wearable devices; Physical activity; Randomized trial.

JEL Classification: I12; I18; I19

*ISER and the Department of Economics and Public Policy, University of Alaska Anchorage, Anchorage, AK 99508. mguettabi@alaska.edu

1 Introduction

New technologies such as wearable sensors with integrated websites are now commercially available and offer consumers a user-friendly tool to monitor their own physical activity performance. The popularity of devices such as Fitbit and similar devices seems to indicate that users enjoy the ability to objectively collect personal physical activity data with a wearable accelerometer, upload these data onto a personal web account, and view summary data to monitor their performance. Hartman et al (2016)[1] make the important point that wearable devices that monitor physically are considerably less expensive than gym memberships and other pricey equipment. While these devices differ in what they measure (heart rate, steps, points, active minutes), their goal is to provide feedback to the user and provide them an interface that monitors their progress. Surprisingly, relatively few have been tested in order to determine their acceptability, usefulness, efficacy or effectiveness in promoting health. An estimated 32 million wearable physical activity trackers will be sold by the end of 2016 and it is projected that sales of these devices will surpass 82 million by 2019 (Parks Associates, 2015; Allied Market Research, 2015). The backdrop of this rise in technology usage is the rapid and continual increase in obesity and its health consequences. Specifically, physical inactivity is a risk factor for many chronic conditions, such as diabetes, cardiovascular disease, and cancer. According to Spear et al (2007)[2], nearly half of preschool children do not meet recommended levels of physical activity (ie, 60 minutes daily) prescribed by the American Academy of Pediatrics. Previous work by Taveras, et al (2007)[3] and Norman, et al (2005)[4] reported barriers to physical activity include a preference for indoor pastimes, low energy levels, time constraints, unsafe neighborhoods, a lack of motivation, not feeling competent or skilled, lack of resources, and insufficient social support from parents and peers. A considerable number of papers have also investigated ways to improve physical activity levels. In most of this literature, children frequently cite fun as the reason they engage in certain activities which means enjoyment is a necessary prerequisite for improving the outcomes of interest. Roberts and Barnard (2005)[5] find that a typical child aged 8 to 10 years spends approximately 65 minutes per day in video game play. These high screen times have been linked to sedentary behaviors and are generally thought to be at least partly responsible for the growing obesity epidemic. Given these high screen times, it may be beneficial to incorporate technology into interventions attempting to encourage improvements in children's physical activity. There is a large literature investigating the role of active and inactive gaming on physical activity amongst them are Leatherdale et al, (2010)[6] and Graf et al, (2009)[7]. There is promising evidence on the beneficial effects of video gaming on improved physical activity outcomes. The extent to which gaming can be incorporated into daily and long term physical activity is yet to be determined. However, it seems clear that providing accessible and appealing options for physical activity in both the school environment and the home will overcome many reported barriers to physical activity. Advances in technology continuously provide researchers with opportunities to develop and test novel intervention modalities that have the potential to deliver effective low cost interventions at the population level.

One particularly interesting line of inquiry is how accelerometers can be used to understand how feedback, incentives, and nudges in general influence objectively measured physical activity. While there has been recent work investigating how incentives affect food choices, there are no papers leveraging this relatively new technology to investigate how incentives affect physical activity in children. This is largely due to the fact that wearable technology is still a relatively new market, with the first commercialized wearable product released in 2008 (Fitbit, 2014). Much of the existing literature surrounding mobile fitness technology focuses on participants that are categorized as overweight and/or are looking to lose weight (see for example Gupta 2011[8], Liu 2011[9]). Our study tries to fill this gap by using an accelerometer, SQORD, designed specifically for kids, which has an accompanying website that allows children to not only check the points earned, but also design their own avatar, see how well their friends are doing, and challenge each other.

1.1 What are SQORDS?

SQORD watches are activity trackers designed to encourage physical activity by having an online platform that allows students to check their scores and those of their peers, see a leaderboard, send messages to their friends, and design an avatar. According to the website (SQORD.com) Sqord is a fun, socially-connected online world powered by real-world play. Players wear the Sqord Booster, which scores the intensity and duration of the physical activity. Move more, get more points, move less, well, you can figure it out. Importantly, SQORD does not produce steps or minutes of vigorous activity. Instead, it produces points which capture the intensity and duration of activity. A step can be roughly translated to 4-6 steps for children.

1.2 What do we know about children's physical activity?

According to the National Survey of Children's Health, 34.2 percent of children aged 6-17 nationally and 28.8 percent of Alaska children exercise 3 days or less. The overweight and or obese rate nationally is 31.3 percent and 29.9 percent for children aged 10 to 17 years old. There is however very little information about activity levels derived from objective metrics for elementary school children.

2 Literature review

According to List and Samek (2015)[11], incentives have also proved to be effective in changing health prevention related behavior in adults. Specifically, they have been shown to be effective for weight loss (Cawley and Price, 2011[12]; Cawley and Price, 2013[13]), and compliance with healthy preventive behaviors (Malotte et al., 1998)[14]. There is, however, debate whether it is either effective or even appropriate to incentivize positive behaviors in children. Opponents of the use of incentives argue that extrinsic rewards crowd out intrinsic motivation and results in outcomes being worse after the end of the incentive period than prior to the introduction of rewards (Deci, Koestner, and Ryan 2001)[15]. However, arguments against the use of incentives sometimes overlook the role that habit formation can play in promoting long run behavioral change defines a habit as "an acquired behavior pattern regularly followed until it has become almost involuntary." If this habit formation process occurs while individuals are incentivized to engage in a behavior, then short-term efforts that encourage children to engage in a particular activity can, if sufficient to overcome any crowding out of intrinsic motivation, result in positive behavior change even after the incentives are removed. Just and Price (2013)[16] provided incentives for five days over a 2-3 week period and found lingering effects during the first two weeks after the intervention, but these did not persist four weeks after the intervention. List and Samek (2015)[11] provided low income school students with a small prize as a reward for choosing a healthier snack (dried fruit) over a less healthy snack (a cookie). They observed a large impact of incentives on the children's choices that persisted even after the incentives were removed, especially when incentives were combined with a health message. On the weight front, Gortmaker et al. (1999)[17] utilized a field experiment to investigate the impact on weight of a 2-year, school-wide educational intervention called Planet Health. They found that Planet Health decreased the prevalence of obesity among girls. Shorter messaging has also been explored.

Our study examines the effect of incentives on children's physical activity over multiple weeks. The design allows us to determine how achievement of objectively measured weekly goals is affected by non-monetary rewards and the role of goal setting.

3 Design: Incentives

In the Fall of 2017, we implemented a goal achievement and incentive program in Anchorage in which children received a wrist band at the end of each week for reaching 70,000 points on at least 3 out of 6 days in a given week. We had 13 schools participate and 7 of them were randomly assigned to the treatment group. The formal experiment lasted form October 14th to December 20th. We, however, have data for five weeks prior to the begining of the experiment which allows to examine pre-intervention trends. All our analysis is done at the week level which gives (week1-week5) as pre-intervention and (week6-week15) as the intervention period. These 13 schools volunteered to participate and were then randomly assigned to implement the incentives/goals or just set the goals and provide feedback at the end of the week.We asked teachers in both groups to allow students to use the SQORD normally without issuing any challenges or competitions during the month of September. Starting October 14th, students in both groups were told by their teachers that they will be setting daily goals at 70,000 daily SQORD points for both boys and girls. These numbers roughly approximate the points that translate to 60 minutes of moderate-vigorous physical activity in a day. The students were also told that on a weekly basis, the target would be considered as achieved if they reached the mark on at least 3 out of 6 days. In addition to these instructions, students in the treatment group were promised rewards if they hit the target. Additionally, we had 29 schools where children had the SQORD device but were not part of the study. This last group allows us to examine the performance of the non-participants relative to those who received incentive+goal target, and those that just received the goal target. We received detailed activity information on points earned for each 15 minute increment.

4 Data

Data protocol: We have a data sharing agreement with the Anchorage School District which allows us access to raw point accumulation for all users. This raw data was sent to us once every two weeks and contained points earned in every 15 minute increment of the day, along with the user's year of birth, and gender. Supplementary materials regarding school characteristics were collected from the Anchorage School District website which provides information on enrollment, ethnicity, and share of students with free or reduced lunch. The characteristics of the three groups of schools(treatment, control, and non study schools) are provided in Table 1. There were 839 students in the treatment group, 838 students in the control group, and 2,299 in the non-study group.

5 Methodology

5.1 Difference in difference

To estimate the average effects of incentives, we estimate two separate equations, both of which offer unique advantages. We compute an indicator variable equal to unity if the school is in the treatment group. The first estimation equation interacts this indicator variable with an indicator for the treatment period. Specifically, we estimate equation (1) below

$$Y_{i,t} = \alpha + \beta(D_i \times \text{Post}_{i,t}) + Z_t + C_i + \epsilon_{i,t}, \tag{1}$$

where $Y_{i,t}$ is the outcome of interest for school i in week t, D_i is the indicator variable identifying the treatment schools, and $Post_t$ is an indicator variable equal to unity for treatment period (week 6 through 15). Any meaningful temporal shocks that are not specific to a single school are captured by time fixed effects Z_t, while any school-specific, time-invariant disturbances are captured by school fixed effects, C_i. The error term, $\epsilon_{i,t}$ is clustered at the school level. Note that the direct effect of $Post_t$ and D_i are both captured by the time and school fixed effects, respectively. Hence, β measures the average effect of being a school in the treatment group from week 6 to week 15, relative to the average effect from week1-week5. This model is well suited to specifically test whether the average treatment effects (the effect of being a school in the treatment group) is statistically different than that in the first half of the sample.

However, a clear concern is that any observed treatment effect is due to pre-existing trend. For example, suppose that the treatment schools throughout gained in the share of students achieving the goal the entire sample period. In this case, β would be positive and significant, but not because of the treatment. To address this concern, we estimate an additional model that allows the treatment effect to vary from one week to another. We specifically estimate equation (2) below:

$$Y_{i,t} = \gamma + \sum_{week1}^{week15} \beta_t (Z_t \times D_i) + Z_t + C_i + \epsilon_{i,t}, \tag{2}$$

where all variables are defined as before. Note that in estimating equation (2), the indicator variable D_i is interacted with week fixed effects and the reference week is week 1. The interpretation of β_t is similar to before, but now it reveals the treatment effect in week t, relative to the treatment effect in week 1. Estimating equation (2) allows us to not only test whether the treatment effect was relatively high at the end of the sample period, but whether the treatment effect rise in tandem with the timing of the intervention.

5.2 Synthetic control method: Investigating heterogeneity

While equations 1 and 2 allow us to assess the effect of incentives and goal setting on all the treated units, there is reason to believe that the intervention we describe above interacts with school characteristics, teacher interest and motivation, other concurrent programs at the school, and general time varying characteristics. This synthetic control analysis we describe below is done at the school level. In other words, we take a multiple case study approach and evaluate the treatment at each school individually. This approach allows us to assess the heterogeneity of the treatment effect. There are a number of advantages of using SCM in this study. First, in program evaluation, researchers often select comparisons on the basis of subjective measures of similarity between the affected and the unaffected regions or states.

But, neither the set of all non-study schools nor a single school likely approximates the most relevant characteristics of the treatment schools (exposed units). SCM, in contrast, provides a comparison school (or synthetic) that is a combination of the schools that did not recieve the treatment — a data-driven procedure that calculates "optimal" weights to be assigned to each borough in the control group based on pre-intervention characteristics — thus making explicit the relative contribution of each control unit to the counterfactual of interest (Abadie and Gardeazabal, 2003; Abadie *et al.*, 2010). SCM provides a systematic way to choose comparison units where the researcher is forced to demonstrate the affinities between the affected and unaffected units using observed characteristics (Abadie *et al.*, 2010[18]; Abadie *et al.*, 2015[19]). We take advantage of this procedure by obtaining a synthetic unit for each of the exposed treatment schools. To be more precise, we analyze each of the treatment schools individually therefore developing a multiple case study approach. This allows us to determine the extent to which the treatment has affected different units. The average effect we describe in the previous section could be driven by one or two schools. By constructing these individual synthetics, we are able to better understand how the treatment interacts with school characteristics and affords us the opportunity to make better recommendations. Our analysis is broken down in three parts:

-First, we use all the schools not receiving the treatment as controls. This set contains both control school which set goals and had feedback as well as schools which did not participate in the study.

-Second, we only use the schools not participating in the study as controls. This set allows us to test the combined effect of goal setting, and incentives.

-Third, we use the study schools not receiving incentives as controls. These schools set goals, and received printouts but did not have an incentive scheme.

An additional advantage as Abadie *et al.* (2010)[18] argue is that unlike the traditional regression-based difference-in-difference model that restricts the effects of the unobservable confounders to be time-invariant so that they can be eliminated by taking time differences, SCM allows the effects of such unobservables to vary with time. In particular, thye show that with a long pre-inervention mathcing on outcomes and characteristics a synthetic control also matches on time-varying unobservables[1].

Finally, because the construction of a synthetic control does not require access to post intervention outcomes, SCM allows us to decide on a study design without knowing its bearing on its findings (Abadie *et al.*, 2010)[18]. The ability to make decisions on research design while remaining blind to how each particular decision affects the conclusions of the study is a safeguard against actions motivated by a "desired" finding (Rubin, 2001)[20].

To obtain the synthetic control we follow Abadie and Gardeazabal (2003)[21] and Abadie *et al.* (2010). For counties $i = 1, ..., J+1$ and periods $t = 1, ..., T$ suppose county $i = 1$ is exposed to the intervention at time $t^* \in (1, T)$. The observed outcome for any state i at time t is

$$Y_{i,t} = Y_{it}^N + \alpha_{it} S_{it}, \tag{3}$$

where Y_{it}^N is the outcome for county i at time t in the absence of the intervention, the binary indicator variable, S_{it}, denotes the intervention taking the value of 1 if $i = 1$ and $t > t^*$, and α_{it}, the coefficient to be estimated, is the effect of the intervention for state i at time t.

Under standard conditions, there exists $\mathbf{W}* = (w_2^*, ..., w_{J+t}^*)'$ such that pre-intervention matching is achieved with respect to the outcome variable as well as characteristics (or predictors), and we can use

$$\hat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}, \quad t \in T_0 + 1, ..., T, \tag{4}$$

[1]As Abadie *et al.* put it, "...only units that are alike in both observed and unobserved determinants of the outcome variable as well as in the effect of those determinants on the outcome variable as well as in the effect of those determinants on the outcome variable should produce similar trajectories of the outcome variable over extended periods time."

as an estimator for α_{1t}. The term $\sum_{j=2}^{J+1} w_j^* Y_{jt}$ on the right-hand-side of (2) is simply the weighted average of the observed outcome of the control counties for $t \in T_o + 1, ..., T$ with weights $\mathbf{W}^*$. The optimal weights placed on each unit are found by minimizing

$$(X_1 - X_0W)'V(X_1 - X_0W), \tag{5}$$

where X_1 is a $k \times 1)$ vector of pre-event predictors for the treatment school, X_0 is a $(K \times J)$ matrix of pre-event predictors for the control group of counties, and W is a $(J \times 1)$ vector of weights that are assigned to controls in the donor pool that sum to one. Finally, V is a $(K \times K)$ diagonal matrix, where the diagonal elements describe the importance of each predictor.

5.3 Inference

Once an optimal weighting vector W* is chosen, the "synthetic" is obtained by calculating the weighted average of the donor pool. The post-intervention values of the synthetic control serve as our counterfactual outcome for the treatment state. The post-intervention gap between the actual outcome and the synthetic outcome, therefore, captures the impact of the intervention. To begin, we follow Bohn et al. (2014)[22] and calculate a difference-in-difference estimate for the treatment state,

$$\Delta_{TR} = |\overline{Y}^{post}_{TR,actual} - \overline{Y}^{post}_{TR,synthetic}| - |\overline{Y}^{pre}_{TR,actual} - \overline{Y}^{pre}_{TR,synthetic}| \tag{6}$$

Where $\overline{Y}^{post}_{TR,actual}$ is the average of the post-intervention actual outcome of the treatment state,$\overline{Y}^{post}_{TR,synthetic}$ is the average of the post-intervention outcome of the counterfactual. Similarly, $\overline{Y}^{pre}_{TR,actual}$ is the average of the pre-intervention actual outcome of the treatment state, and the $\overline{Y}^{pre}_{TR,synthetic}$ is the average of the pre-intervention outcome of the counterfactual. If the outcome changed in response to the intervention in time T_0 we would expect Δ_{TR}>0.

Taking the absolute values in Eq(4) makes sure that the estimate is neutral to the direction of change.

6 Results

6.1 Difference in difference

We begin our analysis using the difference in difference estimation specified in equation(1). In Table 2 column(1), we estimate the effect of being in the treatment group relative to all other schools. To be clear, the "other" group contains both the schools that were in the study but in the control group, and the non-participating schools. The coefficient implies an average increase in the share of students achieving the goal of 70,000 points on at least 3/6 days effect of approximately 10 percentage points during the treatment period. In other words, between week6 and week15, the share of students reaching the desired goal was 10% higher in the treatment schools as a result of the treatment. This increase is only significant at the 10% level. In column (2), we narrow the comparison group to just non-study schools. In this specification, we are estimating the effect of both the incentives and goal setting given that that the non-participating schools had no specific goal setting schemes. We find an average increase in the share of students achieving the goal of 70,000 points on at least 3/6 days effect of approximately 12 percentage points during the treatment period. In other words, between week6 and week15, the share of students reaching the desired goal was 12% higher in the treatment schools as a result of the treatment. This increase is significant at the 5% level. In column (3), we narrow the comparison group to just study schools which were setting goals, receiving feedback, but did receive any incentives for hitting the marks. This comparison between the study schools shows that while the treatment group had a higher share of students achieving the goal (7%), the differences were not statistically significant. One potential concern is that any observed treatment effect is due to pre-existing trend.

For example, suppose that the treatment schools throughout gained in the share of students achieving the goal the entire sample period, that would mean any effects we capture are not necessarily due to the intervention but to other factors. In order to deal with this issue, we estimate equation(2) and present the result in Figure 2(a-c). In Figure 2a, we see that there are no significant differences in the pre-intervention period between the treatment schools and the rest of the schools in our dataset. After week6, the differences are 17% in week 8 but only significant at 10%. The differences become more pronounced starting in week 13 and are largest in weeks 14 and 15. Figure 2(c) shows the dynamic results from the sample using only the non-study schools as controls and the results are very similar to the ones in Figure 2(a) in that the differences become weekly significant in week8 and strongest in weeks 13 and 15. When we narrow the comparison group to just the study school control group, our results are considerably weaker with significant differences in only week 11 and week 13. These results mean that there are no detectable statistical differences between the treatment and control groups at the 5% in the study schools. On average, incentives did not have an independent effect on the share of students achieving the desired goal. We do, however, find that when we compare our treatment schools to non study schools the differences are rather large and statistically significant. This means that goal setting and feedback play a role in determining physical activity in children. All of these results as we explain below, hide the heterogeneity across schools which we address by estimating a multiple case study approach which allows to determine the effect of the treatment on each individual school.

6.2 Investigating heterogeneity: SCM

Estimating an average treatment effect, as we do above, can mask the fact that these treatments are implemented by teachers and are likely sensitive to teacher enthusiasm, classroom competition, and other time varying factors and are likely to have nonuniform effects across schools. Exploring the school variation in this linkage is not only important for a deeper understanding of the treatment but also holds substantive practical ramifications and policy implications. We, therefore, adopt a case-study approach. We use Synthetic Control Method (SCM) for comparative case studies (Abadie and Gardeazabal 2003[21], Abadie, Diamond and Hainmueller 2010[18]) to study treatment school individually instead of aggregating over all treatment schools to estimate the average effect. In particular, we examine the impact of the treatment on the share of students reaching 70,000 points on at least 3 out of 6 days. We find that the treatment did not have uniform effects across schools. Table 3 shows the donor weights for the full sample. This specification is one that uses the largest set of schools and therefore includes non-study participating schools as well as those who participated by were randomly assigned to the non-incentive group.

6.3 Individual treatment schools relative to all other schools

Table 4 and (Figure 2a to Figure 9c) show the individual treatment effect of each school subjected to the intervention relative to the rest of the schools in our dataset. These non intervention donor schools contain control study schools as well as non-study schools. Table 4 includes, the pre-intervention fit, as well as the statistical results of the permutations or randomization tests; i.e., the difference in the post- and pre-intervention mean gap between actual and synthetic outcomes of the treatment units. These gaps are ranked and statistically compared to those from the placebo runs for each of the donor schools. Each school's figures have figure(a) which shows the pre and post-intervention comparisons of actual and synthetics, figure (b) which includes the gap between the actual and synthetic and figure(c) which shows the treatment school's gap relative to the placebos runs from each of the donor schools. We find that 4 out of the 7 treatment schools had a DID rank of 1, one a DID rank of 2. This DID rank of 1 tells us that these 4 treatment schools had the pronouced effect amongst all the units of analysis. For these five schools, there would have had a significantly significant smaller share of students achieving the desired goal of 70,000 points on at least 3 out of 6 days. In the absence of the intervention, these schools' share of students achieving the goal would have been between 10 and 23 percent. These differences tell us that the combination of incentives and goal setting results in better outcomes for 5 out of 7 treated schools relative to the full sample. This sample, however, includes control study schools who have also set goals and therefore were subjected to a portion of the treatment. In the next section (Table 5), we narrow the donor pool to just non-study schools to examine the true combined effect of goal setting and incentives.

6.4 Individual treatment schools relative to non-study schools

Table 5 shows the individual treatment effect of each school subjected to the intervention relative to the non study schools whose students had the SQORD device but did not have a school wide program designed by the research team. This means the donor group all non study schools. Table 4 as table 3 includes, the pre-intervention fit, as well as the statistical results of the permutations or randomization tests; i.e., the difference in the post- and pre-intervention mean gap between actual and synthetic outcomes of the treatment units. These gaps are ranked and statistically compared to those from the placebo runs for each of the donor schools. We find that 4 out of the 7 treatment schools had a DID rank of 1 , one a DID rank of 2. For these five schools, there would have had a significantly significant smaller share of students achieving the desired goal of 70,000 points on at least 3 out of 6 days. In the absence of the intervention, these schools' share of students achieving the goal would have been between 11 and 33 percent. These differences tell us that the combination of incentives and goal setting results in better outcomes for 5 out of 7 treated schools relative to the non-study schools. This effect can be thought of as the incentive+goal setting effect on students achieve This sample, however, includes control study schools who have also set goals and therefore were subjected to a portion of the treatment. In the next section (Table 6), we narrow the donor pool to just study schools to examine the isolated effect of incentives.

6.5 Individual treatment schools relative to control group study schools

Table 6 narrows the analysis to the schools in our randomized control trial. Table 4 shows the individual treatment effect of each school subjected to the intervention relative to the control study schools whose students also were given the 70,000 point target for a minimum of 3 out 6 days but were not given prizes for hitting the target. This means the donor group only includes 6 schools. Table 6 as Table 5 and Table 4 includes, the pre-intervention fit, as well as the statistical results of the permutations or randomization tests; i.e., the difference in the post- and pre-intervention mean gap between actual and synthetic outcomes of the treatment units. These gaps are ranked and statistically compared to those from the placebo runs for each of the donor schools. We find that 5 out of the 7 treatment schools had a DID rank of 1. For these five schools, there would have had a significantly significant smaller share of students achieving the desired goal of 70,000 points on at least 3 out of 6 days. In the absence of the intervention, these schools' share of students achieving the goal would have been between 6 and 25 percent. These differences tell us that incentives resulted in better outcomes for 5 out of 7 treated schools relative to the non-study schools. To be clear, this finding tells us that the treatment schools would have had somewhere between 6 and 25 percent fewer students achieving their goals. This effect can be thought of as the isolated effect of incentives. It appears that incentives were highly effective in motivating students be more active and reach adequate levels of exercise. It is, however, important to note that in two of the schools subjected to the treatment did not perform better that the control schools and therefore there is no evidence of the incentives having been beneficial.

7 Conclusion

We conducted a randomized control trial in the Anchorage school district to determine the effect of goal setting and incentives on children's physical activity. Using a difference in difference estimation, we found that the share of students achieving the desired goal of 70,000 points on at least 3 out of 6 days in the treatment group to be 12% higher than the non-study group during the intervention period. This effect is the sum of both goal setting and incentives. When we include both the control study group and the non study schools, we found that the share of students achieving the goal was 10% higher but only significant at 10%. When we analyzed the isolated effect of incentives and compare the share of students achieving the goal in the treatment group relative to the control study schools, we find that the share of students achieving the goal was 7% higher but not statistically significant. This means, on average, we do not find an isolated effect of incentives on physical activity. When we analyze the effect of incentives and goal settings at the school level, we find that 5 our of 7 schools benefited greatly from both the goal setting and the incentive schemes. Our synthetic control analysis reveals heterogeneities among schools in the effectiveness of the intervention. Specifically, we find that 5 out 7 of the seven schools subjected to the intervention performed considerably better than they would have in the absence of the intervention relative to the control schools who also set goals but had no rewards. Additionally, there are differences in the degree of effectiveness as the range of the effect is between 6 and 25%. We find no detectable effect in two of the schools.

References

[1] Hartman, S.J.,, Nelson, S.H.,Cadmus-Bertman, L.A., Technology and phone based weight loss intervention:Pilot RCT in Women at Elevated Breast Cancer Risk *American J Prev Med* **51:5**, 714–721 (2016).

[2] Spear, BA., Barlow, SE, Ervin,C. Recommendations for treatment of child and adolescent overweight and obesity. *Pediatrics* **120:4**, S254–S288 (2007).

[3] Taveras, E., Field, A., Berkey, S., Rifas-Shiman, S., Frasier, L., Colditz, G., and Gillman, M. Longitudinal Relationship Between Television Viewing and Leisure-Time Physical Activity During Adolescence. *Pediatrics* **119:2**, (2005).

[4] Norman, GJ, Schmid, BA, Sallis, JF, Clafas, KJ, and Patrick, K. Psychosocial and environmental correlates of adolescent sedentary behaviors *Pediatrics* **116:4**, 908–916 (2005).

[5] Roberts, C, Barnard, J. Effects of exercise and diet on chronic disease *Journal of Applied Physiology* **98:1**, 3-30 (2005).

[6] Leatherdale, ST., Woodruff, SJ., Manske, SR. Energy expenditure while playing active and inactive video games *Am J Health Behav* **34:1**, 31–35 (2010).

[7] Graf, DL, Pratt, LV, Hester, CN, Short, KR. Playing active video games increases energy expenditure in children *Pediatrics* **124:2**, 534–540 (2009).

[8] Gupta, N. Digital Fitness Connector: Smart Wearable System. *IEEE Computer Society* , 118-121 (2011).

[9] Liu, C, Qing, Z, Holroyd, K, Seng, E. Status and Trends of Mobile-health Applications for IOS Devices: A Developer's Perspective. *Journal of Systems and Software* **84:11**, 2022-2033 (2011).

[10] Collier, P, Goderis B, Commodity Prices, Growth, and the Natural Resource Curse: Reconciling a Conundrum *Centre for the Study of African Economies Working Paper Series* , (2008).

[11] List, JA., Samek, AS, The behavioralist as nutritionist: Leveraging behavioral economics

to improve child food choice and consumption. *Journal of Health Economics* **39**, 133–146 (2014).

[12] Deci, E, Cawley, John, Price, Joshua, Outcomes in a Program that Offers Financial Rewards for Weight Loss *Chapter 4 in Economic Aspects of Obesity, ed. Grossman Michael, Mocan Naci, 91-126. Chicago Ill: NBER and University of Chicago Press.* **4**, 91-126 (2011).

[13] Deci, E, Cawley, John, Price, Joshua, A case study of a workplace wellness program that offers financial incentives for weight loss *Journal of Health Economics* **32**, 794-803 (2013).

[14] Malotte, C, Rhodes, F, Mais, K, Extrinsic rewards and intrinsic motivations in education: Reconsidered once again. *American Journal of Public Health* **88**, 792-796 (1998).

[15] Deci, E, Koestner, R, Ryan, R, Tuberculosis Screening and Compliance with Return for Skin Test Reading among Active Drug Users *Review of Educational Research* **71(1)**, 1-27 (2001).

[16] Just, D, Price, J, Using incentives to encourage healthy eating in children *Journal of Human Resources* **48(4)**, 855-872 (2013).

[17] Gortmaker, SL, Peterson, K, Wiecha, J Reducing obesity via a school-based interdisciplinary intervention among youth *Arch Pediar Adolesc Med* **153**, 409-418 (1999).

[18] Abadie, A, Gardeazabal, J, Synthetic control methods for comparative case studies: estimating the effect of California's Tobacco Control Program. *J Am Stat Assoc* **105(490)**, 493-505 (2010).

[19] Abadie, A, Diamond, A, Hainmueller, J Comparative Politics and the Synthetic Control Method *American Journal of Political Science. 2014* **59**, 495-510 (2015).

[20] Rubin, Donald, Using propensity scores to help design observational studies: Application to the tobacco litigation *Health Services & Outcomes Research Methodology* **2**, 169-188 (2001).

[21] Abadie, A, Gardeazabal, J, The Economic Costs of Conflict: A Case Study of the Basque Country *American Economic Review* **93(1)**, 113-132 (2003).

[22] Bohn, S, Lofstrom, M, Raphael, S Did the 2007 legal arizona workers act reduce the state's unauthorized immigrant population? *The Review of Economics and Statistics* **96(2)**, 258-269 (2014).

Table 1: Summary statistics

Variable	Treatment Mean	Treatment S.D	Control Mean	Control S.D	Non-study Mean	Non-study S.D
Number of students	839	-	838	-	2,299	-
Share of students $\geq$ 70k 3 days a week	0.378	0.120	0.298	0.114	0.290	0.143
Daily activity	51,331	44,780	46,616	42,245	45,563	41,794
Percent with free or reduced lunch	0.284	0.137	0.386	0.170	0.395	0.294
Percent white	0.595	0.139	0.434	0.194	0.448	0.230
Percent native	0.061	0.033	0.090	0.027	0.055	0.047
Percent male	0.515	0.499	0.504	0.50	0.502	0.500

Graph 1: Share of students achieving 70,000 on at least 3 out of 6 days

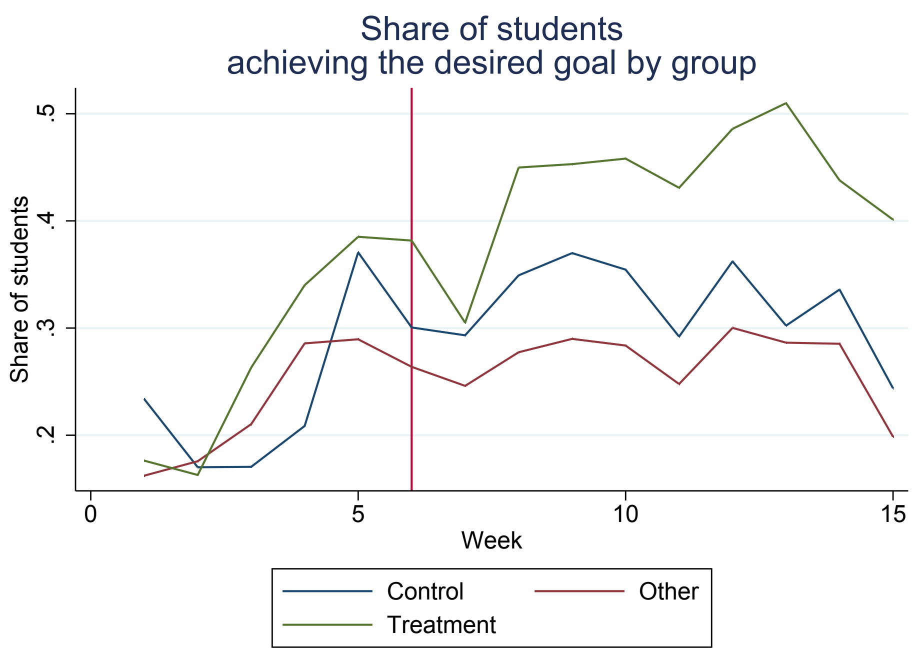


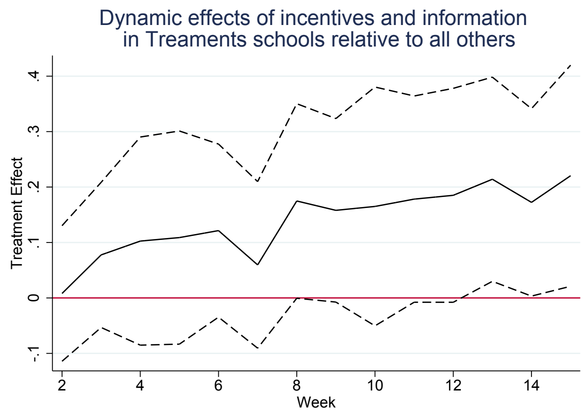
Table 2: Difference in difference results: Equation 1

	Relative to everybody else	Just non-study schools	Relative to control groups
$D_i \times \text{Post}_t$	0.106*	0.123***	0.0762
	(0.0567)	(0.0574)	(0.0625)
R^2	.496	.565	.517
N	510	405	195

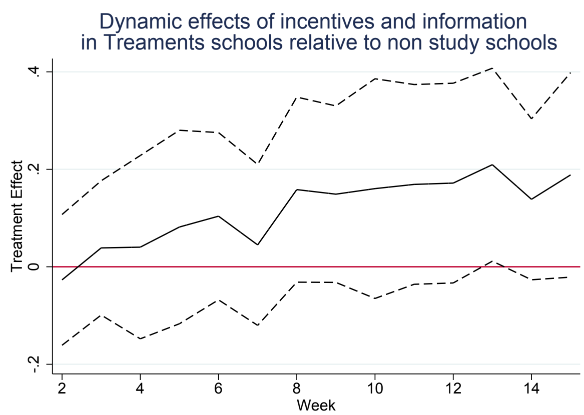
Note. ***, **, * corresponds to 1%, 5% and 10% significance, respectively. The control groups are shown in the column headers. Standard errors (clustered at the school level) are given in parenthesis below the estimated coefficients. Year and school fixed effects are included in all regressions. The coefficients can be interpreted as the causal effect of the treatment on the share of students achieving 70,000 points on at least 3 out of 6 days.

Share of students achieving 70,000 on at least 3 out of 6 days

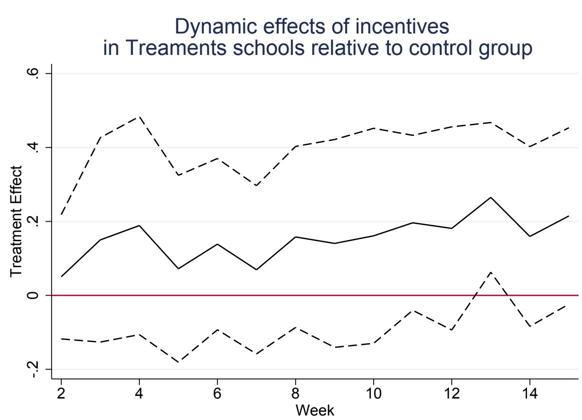
Graph 2a: Treatment relative to all other schools



Graph 2b: Treatment relative to non-study schools



Graph 2c:Treatment relative to control study schools



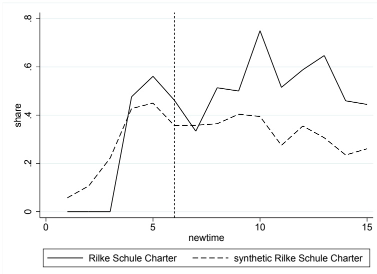
Note: 95 percent confidence intervals are given. The solid line in each diagram describes the weekly treatment effect estimated from equation 2

Table 3: Donor weights for the full sample : Relative to all schools

Organization id	**Organization name**	Rilke	Gladys	Chugiak	Chugach	Bowman	Ursa	Turnagain
324	Abbott Loop	-	-	-	-	-	-	
587	Alpenglow	-	-					
588	Aquarian	0.409	-	-		0.052	-	0.197
325	Baxter	0.273	-	-	-	-	-	
326	Bear Valley	-	-	-	-	-	-	0.23
327	Birchwood ABC	-	-	-	-	-	0.163	
328	Campbell	-	-	0.839	-	0.328	-	0.51
329	Chinook	-	-	-	-	-	-	
332	College Gate	-	-					
591	Creekside Park	-	-	-	-	-	-	
333	Denali	-	0.131	-	-	-	0.252	
784	Eagle Academy	-	-	-	-	-	-	
622	Eagle River	-	-	-	-	-	-	
335	Fire Lake	-	-					
846	Gilson Middle	-	-	-			-	
590	Girdwood	-						
598	Government Hill	-	-	-	-	-	-	
240	Hermon Hutchen	-	-	-	-	-	-	
337	Homestead	-	-	-	-	0.008	-	
338	Huffman	-	-	0.161	0.002	0.528	-	
339	Inlet View	-	-	-		-		
340	Kasuun	0.318	-	-	-	-	-	
341	Kincaid	-					-	
342	Klatt	-	-	-	-	-	-	
343	Lake Hood	-	-	-	-	-	-	
344	Lake Otis	-	-	-			-	
498	Main Elementary	-	-	-			-	
592	Mt Spurr	-	-	-	-	-	-	
248	Muldoon	-	0.45	-	-	-	-	
345	North Star	-	-	-	-	-	-	
346	Northern Lights ABC	-	-	-	-	-	-	
347	Northwood	-	-	-	-	-	-	
348	Nunaka Valley	-	0.419	-	-	-	0.417	
593	O'Malley	-	-	-	0.8	-	-	
349	Ocean View	-	-	-	-	-	-	
247	Orion	-	-	-	-	-	0.13	
623	Peterson	-	-	-	-	-	-	
594	Polaris	-	-	-	-	-	-	
785	Rabbit Creek	-	-	-		-		
351	Ravenwood	-	-	-	-	-	0.039	
352	Rogers Park	-	-	-	-	-	-	
596	Sand Lake	-	-	-	-	-	-	
250	Scenic Park	-	-	-	-	-	-	
469	Seward Middle	-	-	-	-	-	-	
353	Spring Hill	-	-	-	-	-	-	
354	Susitna	-	-	-	-	-	-	
244	Tudor	-	-	-	-	-	-	
787	Tyson	-	-	-	-	-	-	
468	Williwaw	-	-	-	0.198	0.084	-	0.063
357	Willow Crest	-	-	-	-	-	-	

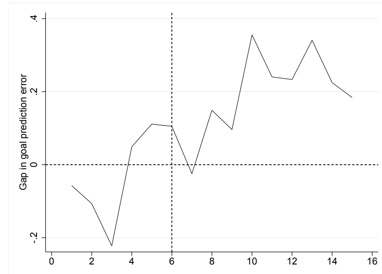
Rilke:

Graph 3a: Actual and synthetic share of students achieving the goal



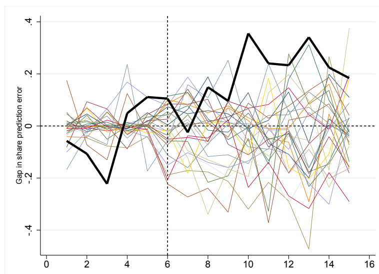
Note:

Graph 3b: Actual minus synthetic



Note:

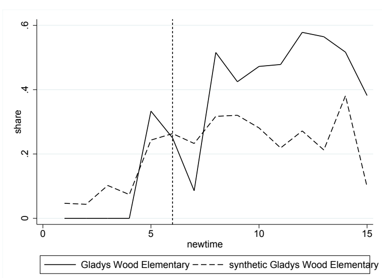
Graph 3c: Placebos



Note:

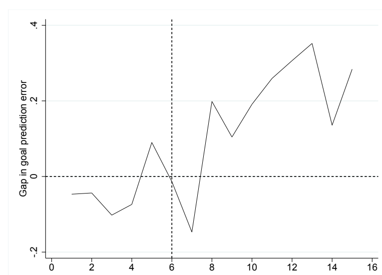
Gladys

Graph 4a: Actual and synthetic share of students achieving the goal



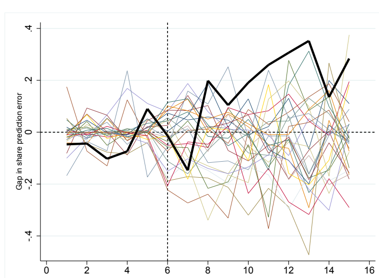
Note:

Graph 4b: Actual minus synthetic Per capita income maintenance



Note:

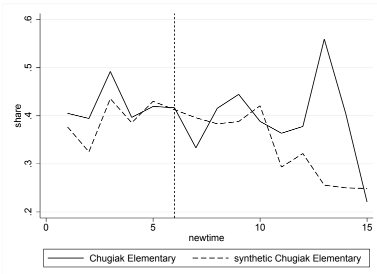
Graph 4c: Placebos Per capita income maintenance



Note:

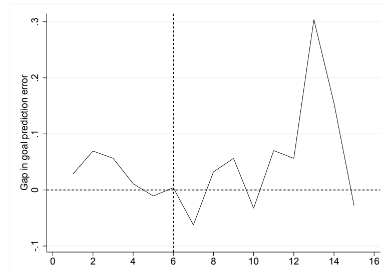
Chugiak

: Graph 5a: Actual and synthetic Chugiak



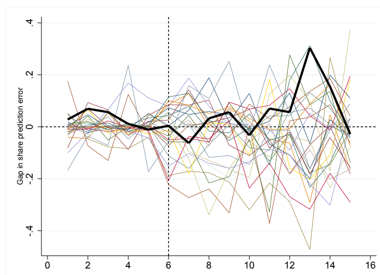
Note:

Graph 5b: Actual minus synthetic



Note:

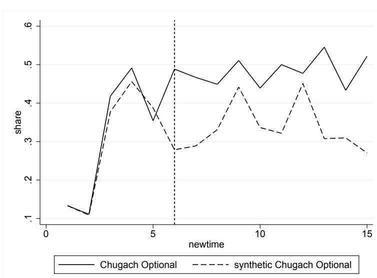
Graph 5c: Placebos



Note:

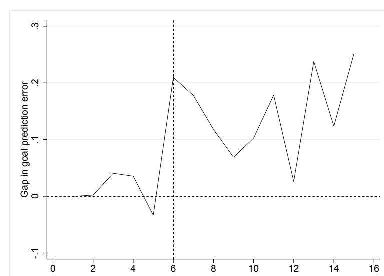
Chugach

: Graph 6a: Actual and synthetic share of students achieving the desired goal



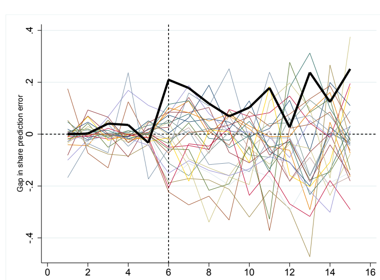
Note:

: Graph 6b: Actual minus synthetic



Note:

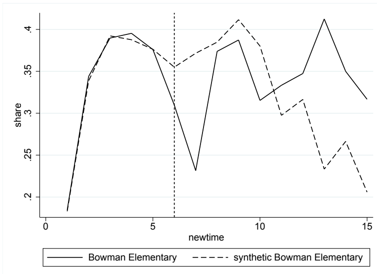
: Graph 6c: Placebos



Note:

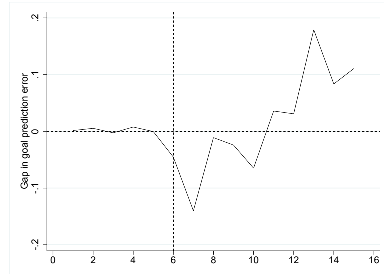
Bowman

: Graph 7a: Actual and synthetic share of students achieving the desired goal



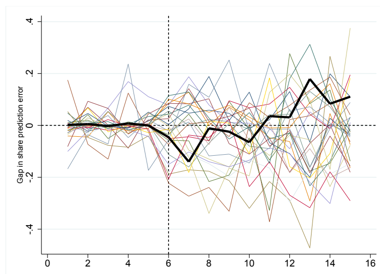
Note:

: Graph 7b: Actual minus synthetic



Note:

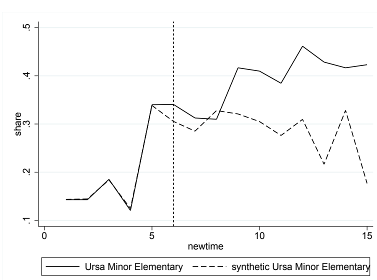
: Graph 7c: Placebos



Note:

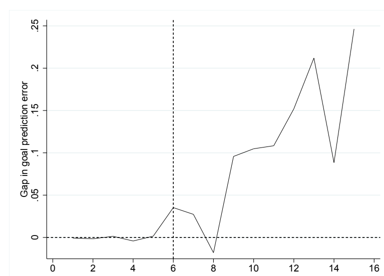
Ursa

: Graph 8a: Actual and synthetic share of students achieving the desired goal



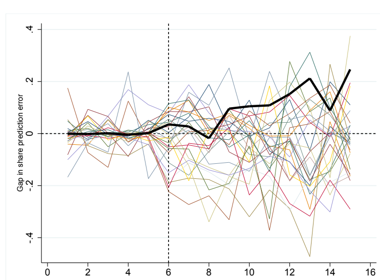
Note:

: Graph 8b: Actual minus synthetic



Note:

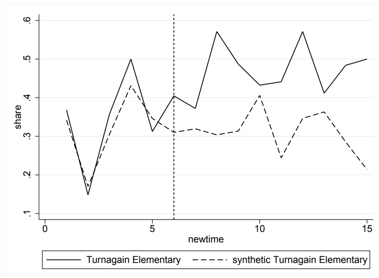
Graph 8c: Placebos



Note:

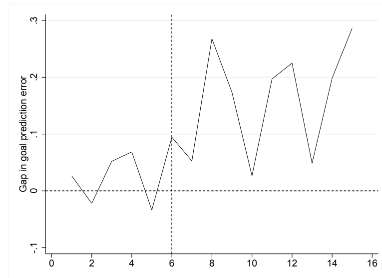
Turnagain

Graph 9a:Actual and synthetic



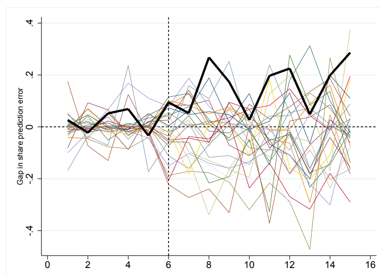
Note:

Graph 9b: Actual minus synthetic



Note:

Graph 9c: Placebos



Note:

Table 4: Estimation statistics (Full sample)

	Rilke	Gladys	Chugiak	Chugach	Bowman	Ursa	Turnagain
SCM:pre-intervention fit							
Absolute prediction error/mean ratio	0.192	0.55	0.073	0.0296	0.006	0.0038	0.054
SCM inference:permutations test							
Pre-intervention difference (D1)	-0.0452	-0.0352	0.030	0.0089	0.0023	-0.00072	0.0182
Post-intervention difference (D2)	0.1904	0.167	0.055	0.1494	0.0154	0.105	0.157
DID=$\|D2\| - \|D1\|$	0.235	0.203	0.024	0.1404	0.0131	0.10593	0.138
P-value:DID	0.037	0.037	0.296	0.037	0.296	0.074	0.037
DID rank	1	1	8	1	8	2	1

Table 5: Estimation statistics (Relative to non-study schools)

	Rilke	Gladys	Chugiak	Chugach	Bowman	Ursa	Turnagain
SCM:pre-intervention fit							
Absolute prediction error/mean ratio		-	-	-	-		
SCM inference:permutations-test							
Pre-intervention difference (D1)	-0.0383	-0.0306	0.0312	-0.00577	0.002	-0.00592	0.01836
Post-intervention difference (D2)	0.2033	0.1834	0.0556	0.3312	0.0154	0.109	0.1569
DID=$\|D2\| - \|D1\|$	0.241	0.214	0.0244	0.3370	0.0132	0.115	0.1385
P-value:DID	0.0476	0.0476	0.285	0.047	0.285	0.095	0.0476
DID rank	1	1	6	1	6	2	1

Table 6: Estimation statistics (Relative to control schools)

	Rilke	Gladys	Chugiak	Chugach	Bowman	Ursa	Turnagain
SCM:pre-intervention fit							
Absolute prediction error/mean ratio		-	-	-	-		
SCM inference:permutations-test							
Pre-intervention difference (D1)	-0.010	-0.094	0.108	-0.00577	0.079	-0.0149	0.046
Post-intervention difference (D2)	0.248	0.093	0.064	0.3312	0.0207	0.0628	0.1082
DID=$\|D2\| - \|D1\|$	0.258	0.187	-0.0475	0.337	-0.0584	0.0778	0.061
P-value:DID	0.142	0.142	0.857	0.142	0.857	0.142	0.142
DID rank	1	1	6	1	6	1	1

Table 7: What do the results mean?
Actual minus synthetic

	Rilke	Gladys	Chugiak	Chugach	Bowman	Ursa	Turnagain
Average effect:							
Week 1-Week 15	0.111	0.099	0.047	0.102	0.011	0.069	0.110
Dynamic effect							
Week 1	-0.057	-0.046	0.028	-0.000	0.001	-0.000	0.026
Week 2	-0.106	-0.043	0.069	0.002	0.005	-0.001	-0.022
Week 3	-0.222	-.102	0.056	0.04	-0.002	0.001	0.052
Week 4	0.049	-0.073	0.111	0.035	0.007	-0.004	0.068
Week 5	0.111	0.090	-0.0107	-0.033	-0.000	0.0015	-0.033
Week 6	0.105	-0.0134	0.003	0.209	-0.045	0.035	0.094
Week 7	-0.024	-0.147	-0.062	0.177	-0.140	0.0273	0.052
Week 8	0.149	0.198	0.032	0.117	-0.011	-0.018	0.267
Week 9	0.0961	0.104	0.056	0.068	-0.024	0.095	0.173
Week 10	0.355	0.191	-0.032	0.102	–0.064	0.104	0.026
Week 11	0.240	0.259	0.070	0.178	0.035	0.108	0.197
Week 12	0.233	0.306	0.056	0.026	0.030	0.151	0.224
Week 13	0.340	0.352	0.303	0.237	0.179	0.212	0.048
Week 14	0.224	0.135	0.154	0.123	0.083	0.088	0.198
Week 15	0.183	0.283	-0.027	0.251	0.110	0.246	0.286