

Outcomes in a Program that Offers Financial Rewards for Weight Loss¹

John Cawley

Department of Policy Analysis and Management, Cornell University
and National Bureau of Economic Research

Joshua A. Price

Department of Policy Analysis and Management, Cornell University

November 6, 2008

Paper to be presented at the NBER Conference on Economic Aspects of Obesity,
November 10-11, 2008 at Louisiana State University

Abstract

Obesity rates in the U.S. have doubled since 1980. Given the medical, social, and financial costs of obesity, a large percentage of Americans are attempting to lose weight at any given time but the vast majority of weight loss attempts fail. Researchers continue to search for safe and effective methods of weight loss, and this paper evaluates one promising method - offering financial rewards for weight loss.

This paper studies data on 2,351 employees in 15 worksites who participated in a year-long worksite health promotion program that offered financial rewards for weight loss. The intervention varied by employer, in some cases offering steady quarterly rewards and in other cases requiring participants to post a bond that would be refunded at year's end conditional on weight loss. Still others received no financial incentives at all and serve as a control group.

This paper documents far higher attrition in this real-world intervention than has been experienced in similar interventions designed and operated by researchers. Moreover, average weight loss due to the financial rewards is small: after one year, it is 1.7 pounds for those who posted a refundable bond and is not significantly different from zero for those facing steady quarterly rewards. We conclude by offering concrete suggestions for increasing the effectiveness of such programs.

¹ The authors thank Company X for providing their data and for the generosity of their time in explaining their intervention and discussing the data. The authors thank Dan Benjamin, Ron Ehrenberg, participants in a Cornell work-in-progress seminar, and participants in the NBER pre-conference on the Economics of Obesity for helpful comments.

Introduction:

Obesity, defined as a body mass index² (BMI) of 30 or higher, has more than doubled in U.S. since 1980 (Flegal et al. 2002). As of 2003-2004, 66.3 percent of Americans were at least overweight (BMI of 25 or higher) and 32.2 percent were obese (Ogden et al. 2006).

Obesity imposes a variety of health, financial, and psychic costs: greater risk of illnesses such as diabetes and hypertension (Must et al. 2006), higher out-of-pocket health care costs (Finkelstein, Fiebelkorn, and Wang 2003), higher job absenteeism (Cawley et al. 2007), lower quality of life, especially for women (Muennig et al. 2006), lower wages for women, especially white women (Cawley 2004), and greater risk of death, especially from cardiovascular disease (Flegal et al. 2007).

A variety of approaches are being used to treat obesity and encourage weight loss. One promising strategy based on behavioral economics is to offer financial incentives for weight loss. Obesity is costly to health insurance companies (Finkelstein et al. 2003) and employers (Cawley et al. 2007), so for either or both of those organizations to offer monetary incentives for enrollees or employees to lose weight could be mutually beneficial.

This paper studies data from a firm that coordinates a program of financial incentives for weight loss in various worksites in the U.S. We study attrition and weight loss in two types of incentive programs: one that offers quarterly payments that rise in value with the amount of weight loss, and a second that has two features: it takes deposits (bonds) that are only refunded if the employee achieves a specific weight loss goal and

² BMI is defined as weight in kilograms divided by height in meters squared.

also includes a quarterly lottery for those who have lost weight. Relative to previous studies of weight loss in response to financial incentives, our study offers a relatively large sample size (2,351) and relatively long intervention (1 year).

A 2006 Institute of Medicine report on obesity prevention set the immediate next step – which it described as an essential priority action for the near future – as “learning what works and what does not work and broadly sharing that information.” (IOM 2006, p. 410). It also notes that “All types of evaluation can make an important contribution to the evidence base upon which to design policies, programs, and interventions.” (Ibid, p. 4). This paper makes a contribution to that effort by documenting the weight loss and attrition in one promising anti-obesity program. We conclude with suggestions for improving the design and effectiveness of such programs.

Conceptual Framework and Previous Literature

For obese people, weight loss would likely result in substantial benefits.³ For example, the health benefits of modest weight loss (defined as 5-10% of starting weight) include decreased blood pressure and cholesterol and a 25% reduction in mortality risk for type 2 diabetics (Vidal, 2002). The obese also report lower quality of life (Ford et al. 2001). There also may be financial benefits of weight loss for the obese. Cawley (2004) finds a causal impact of weight on wages, and that obese white females earn roughly 11 percent less than healthy-weight white females. Finkelstein et al. (2003) calculate that,

³ There are two ways researchers have sought to measure the benefits of weight loss. The first is to examine changes in outcomes associated with losing weight. The second is to compare the outcomes of individuals of different weight, and assume that the difference in outcomes is due to the difference in weight. Each has its limitations: weight loss studies often lack power and comparisons across weight levels are confounded by differences in unobserved characteristics. Vidal (2002) assesses the evidence on the benefits of weight loss and concludes that modest weight loss (5-10% of initial body weight) improves cardiovascular risk factors and helps prevent or delay the onset of type 2 diabetes and hypertension.

relative to the healthy weight, the obese incur \$125 higher annual out-of-pocket health care costs. With two-thirds of Americans are overweight or obese (Ogden et al. 2006), and given these potential benefits of weight loss, it may not be surprising that 46 percent of all American women and 33 percent of all American men are trying to lose weight (Bish et al. 2005).

Most people fail in their attempts to lose weight⁴, and many of those who are successful in losing weight regain it in a short period of time.⁵ For example, in one community-based study of weight gain prevention (Crawford et al., 2000), most (53.7%) participants *gained* weight in the first twelve months, three-quarters gained weight over three years, and only 4.6% lost weight and maintained the loss for three years.

Theory and evidence from psychology and behavioral economics provide several explanations for why so many weight loss attempts fail. First, the benefits of weight loss are not salient. For example, foregone quality of life and lost wages are not visible and therefore they are frequently unrecognized as opportunity costs (Bastiat, 1850) or as possible benefits of weight loss.

A second possible explanation for repeated failure at weight loss is that the benefits of weight loss may not be immediate. Improvements in health and labor market outcomes may not occur for some time after weight loss, and Ainslie (1975) finds

⁴ Some obese individuals are able to lose weight by modifying their behaviors: eating less and exercising more. In the select group enrolled in the Weight Control Registry, all of whom have lost at least 30 pounds and kept it off for at least one year, 44.6% report losing the weight entirely on their own, that is, without the help of a commercial program, physician, or nutritionist (Wing and Phelan, 2005). Clearly such statistics do not generalize to the population; anyone who failed at initial weight loss is ineligible for this registry of people who maintained weight loss for a year.

⁵ Conventional wisdom is that virtually no one succeeds at maintaining weight loss. This perception has been traced back to a 1959 study of 100 obese individuals in which only 2 percent maintained loss of 20 pounds or more two years after the treatment (Stunkard and McLaren-Hume 1959; Wing and Phelan 2005). However, the 1959 study was based on a crude diet intervention with negligible support or follow-up so its poor results may not generalize to today's much more intensive interventions.

consistent evidence that there is a decline in the effectiveness of rewards as the rewards are delayed from the time of choice.

A third explanation for repeated failure at weight loss is that, contrary to the standard economic model of discounted utility (Samuelson 1937), people may discount hyperbolically, generating time-inconsistent preferences (Ainsley 1975). In this context, time inconsistent preferences mean that people want to do what is in their long-run interest (lose weight), but they consistently succumb to the temptation to eat and be sedentary. Thaler and Shefrin (1981) describe individual decision-making as a battle between a farsighted planner (who in this context wants to diet) and a myopic doer (who in this context wants to eat and be sedentary).

One intervention, financial rewards for weight loss, may offer a solution to the problems of salience, immediacy, and time-inconsistency. Financial rewards, even though they may be dwarfed in value by the other benefits of weight loss, have the benefit of being salient, with their amount and delivery date known with certainty in exchange for clearly defined objectives. Even small financial incentives can be effective because research has found that people tend not to compare payoffs to their income or wealth but instead “bracket” them - consider them in isolation (Read et al. 1999; Kahneman and Tversky 1979). Lotteries may be particularly cost-effective incentives for healthy behavior. People tend to overweight the probability of unlikely events and underweight the probability of likely events (Kahneman and Tversky 1979), implying that lotteries can be more attractive than certain payments even if the two have equal expected values. Financial rewards can also be paid immediately, before other benefits of weight reduction may be realized.

Financial rewards can also be structured to help people with time-inconsistent preferences stay committed to weight loss. In general, pre-commitment devices may help people with time-inconsistent preferences empower their farsighted planner (Strotz 1956; Laibson 1997). In this context, one could allow people to post a bond that is automatically forfeited if they fail to achieve their weight loss goals. Such a bond allows a person to influence their own future decisions by increasing the punishment for succumbing to short-run temptation. People tend to exhibit loss aversion – they dislike losing their own money more than they like winning an equal amount of someone else’s money (Tversky and Kahneman 1991; Camerer 2005), which suggests that a posted bond may be more effective than a reward of the same size. Using a bond to increase adherence to a weight-loss regimen does not guarantee success. Even individuals who are aware of their time-inconsistent preferences may still be partially naive in that they overestimate their future willpower (O’Donoghue and Rabin 2001) and as a result may either post too small a bond or have too much faith in the bond as a precommitment device.

Motivated by these theories and findings, several businesses now help employers offer financial incentives for employee weight loss: e.g. Tangerine Wellness, IncentaHealth, and Wellfolio. In addition, several businesses help consumers post bonds that are only refunded if one achieves specific weight loss goals. The William Hill betting agency in the U.K. books wagers that the bettor cannot achieve a specified weight loss in a specific period of time (Burger and Lynham 2008).⁶ William Hill verifies the

⁶ This market is relatively small – the annual number of applications for such bets is roughly 200 (Burger and Lynham, 2008)

weight loss with medical exams. A new company named stickK.com⁷ that was founded by Yale economists Ian Ayres and Dean Karlan allows people to post bonds that are forfeited if they fail to meet their weight loss goal. However, verification is weak: success in achieving one's goal is determined (and refunds are made) based on either the honor system or through verification by a third party chosen by the bettor, and if the third party doesn't submit a report the self-report of the bettor is accepted.

The contribution of this paper is to examine outcomes in a program that offers various financial rewards (including certain payments, lotteries, and refundable bonds) for weight loss. The outcomes we examine include attrition and weight loss, both in pounds and as a percentage of baseline weight. We also examine whether the data are consistent with specific hypotheses about who is likely to attrite and lose more weight.

A substantial literature confirms that financial incentives influence healthy behaviors. Kane et al. (2004) review 42 studies of the effect of economic incentives on preventive behaviors such as immunization, smoking cessation, and exercise; they find that the economic incentives were effective at changing behavior in 73% of studies. Financial incentives form the basis for an innovative substance abuse treatment program known as contingency management. A meta-analysis found overwhelming evidence that such incentives raise compliance (drug abstinence) by an average of 30 percent (Lussier et al., 2006). Consistent with bracketing, even small financial incentives have proven effective; for example, as little as \$2.50 for a single negative test result for cocaine (Higgins et al., 2002).

⁷ The website's Frequently Asked Questions page states that the company's name includes two K's because "K" often symbolizes "contract" in legal writing.

Specific to the current context, there is mixed empirical evidence on the extent to which weight loss is responsive to financial rewards. A recent review and meta-analysis (Paul-Ebhohimhen and Avenell 2007) identified nine published randomized controlled trials (RCTs) that used guaranteed financial incentives (i.e. certain payments, not lotteries) for weight loss, with a follow-up of at least one year. The meta-analysis was unable to reject the null of no effect of financial rewards on weight loss; it calculated a mean weight loss of 0.4 kg at 12 months, which was not statistically significant. A broader set of studies (e.g. with non-randomized designs or shorter follow-up) are listed in Appendix Table 1.⁸

Relative to past studies, ours has several advantages. This study has a relatively large sample size (2,351); for comparison, the sample size of all published RCTs of financial incentives for weight loss *combined* totals 424 (treatment N=252, control N=172) (Paul-Ebhohimhen and Avenell, 2007). The greater sample size will give us greater power to detect effects. The intervention studied by this paper also covers a relatively long time period (one year). Moreover, we examine data from a real-world intervention rather than one constructed by and overseen by researchers, which is important because a criticism of studies of weight loss programs is that it is unclear how the results of pilot programs generalize to real-world implementation.

Description of the Intervention

Our data come from a company (that we will call Company X) that helps employers provide financial incentives for their employees to lose weight; specifically, it

⁸ There are other studies that offer financial rewards for exercise or for attending weight loss programs, but Appendix Table 1 is limited to studies of financial rewards for weight loss.

monitors employee weight loss and pays the rewards. After an employer contracts with Company X, Company X has a kickoff event in the workplace that explains the program to the employees and encourages them to sign up. Participation is optional. Those who sign up select a physical activity regimen at either the foundation (easiest), intermediate, or advanced level. The program consists of several elements: 1) daily email coaching that includes information about healthy and effective methods of weight loss including decreasing calorie intake and increasing physical activity in a manner consistent with the regimen the enrollee chose at baseline; 2) call center support; 3) weigh-ins at least once a quarter; and 4) financial incentives for achieving specific weight loss targets. Only employees who are overweight (BMI of at least 25) are eligible to receive financial rewards, and no financial rewards will be paid once an employee's BMI falls below 25 (i.e. when the employee falls into the "healthy weight" category).

The weigh-ins take place in HIPAA-compliant kiosks that company X installs in the employer's workplace. Employees enter the privacy-protected kiosk and stand on a scale; their body mass index is recorded and sent over an internet connection to their personal webpage as well as to Company X's database. Participants can weigh themselves as often as they like, and the lowest recorded weight will be counted as that quarter's weight. Percent of baseline weight lost determines the financial rewards.⁹

Company X has a standard set of incentives that it proposes, but employers can modify it. In our data, there are three incentive schedules. The first is Company X's

⁹ We asked Company X whether people game the system by trying to weigh more at baseline (from which future weight losses are judged). They said that from the cameras in the kiosks they do not see people wearing heavier clothes to the baseline weigh-in than to later weigh-ins; in all cases people seem for vanity reasons to remove shoes and sweaters before weighing in. However, Company X acknowledges that they have no way to know if people (e.g.) hid weights in their pockets or shoes before the baseline weigh-in. If people engage in such deception then we would expect to see significant drops in weight at the first weigh-in after baseline but we do not find this pattern in the data.

standard set of incentives: the employee participants pay no fee (all costs are paid by the employer), and employees receive quarterly payments determined by their progress toward specific weight loss goals. Table 1 lists the standard set of incentives: payment thresholds occur at each percentage point of weight loss up to 5% (1, 2, 3, 4, 5), then thresholds occur every 5 percentage points (5, 10, 15, 20, 25, 30) up to 30% of weight loss. The payment associated with these thresholds varies; for the first seven (1, 2, 3, 4, 5, 10, 15) the reward is a dollar per percentage point of weight loss. Then the percentage-point rewards increase: \$25 for losing 20%, \$35 for losing 25%, and \$50 for losing 30%. These are monthly amounts that are paid quarterly, so someone who loses 5% of his weight and keeps it off for three months receives a \$15 check for the quarter (\$5 monthly payment x 3 months). Five employers (with a total of eleven worksites participating) used this standard incentives schedule.

The second incentive schedule, used by one employer (with two worksites participating), is shown in Table 2 and includes both a lottery and a deposit contract. The lottery takes place each quarter and the prizes are gift certificates (ten \$50 gift cards and ten \$50 salon vouchers); only those who had lost some weight since baseline are eligible for the drawing. The deposit contract is that employees must pay \$9.95 per month, all of which ($12 * \$9.95$ or \$119.40) is refunded at the end of the year if the respondent loses at least 5% of baseline weight by year's end. If the respondent loses 10% or more of their baseline weight, they receive in addition to their refunded fees (\$119.40) a \$100 bonus, for a total of \$229.40. In addition, the "biggest loser" (as a percent of baseline weight) receives a \$250 gift certificate at the end of the year.

Whether a participant would receive a higher payoff in the standard or modified group depends on both quarter and magnitude of weight loss. In quarters one through three, the standard incentives are more generous than the modified incentives in quarters one through three at all levels of weight loss, with the exception that those losing between 0.1% and 0.9% of baseline weight receive no reward in the standard incentives group but are eligible for the lottery for gift cards in the modified incentives group. In quarter four, the standard incentives are more generous for weight loss of between 1% and 4%, but the modified incentives are more generous for weight loss of 5% or more.

The third schedule, used by one employer (with a total of two worksites), offered no incentives for weight loss, but did include one small incentive to not attrite: participants were promised \$20 if they participated for the entire year (i.e. weighed in at least once in each of the four quarters). This group received all of the features of the Company X intervention (daily emails, call center access, weigh-ins at the kiosk) but had no incentives for weight loss, making it useful both as a control group for measuring the impact of financial incentives isolated from all the other program elements, and for estimating the impact of the Company X treatment minus the financial incentives.

Figure 1 presents a flow diagram of attrition and analysis for all three groups (standard incentives, modified incentives, control) combined.

Hypotheses

Part of our purpose in this paper is exploratory - to measure enrollment, attrition and weight loss in these programs. We focus in particular on attrition and weight loss as outcomes because the NIH Technology Assessment Conference Panel (1993)

recommends using the percentage of all beginning participants who complete the program, and the percentage of those completing the program who achieve various degrees of weight loss as measures of program success. The NIH considers a loss of 10% of baseline weight in 6 months to one year to be good progress for an obese individual (NHLBI 2000).

Another purpose of this paper is to test the following hypotheses.

Hypotheses Regarding Enrollment

There will be lower enrollment in the program that required people to post forfeitable bonds. The law of demand states that the quantity demanded falls with price. The program that requires people to post a forfeitable bond raises the expected price of participation, assuming that not all possible participants expect a 100% probability of success (and therefore the return of their bond). The published literature confirms that, all else equal, enrollment in weight loss programs is lower if people are asked to post forfeitable bonds (e.g. Jeffery et al. 1978).

Those who are willing to post a forfeitable bond (i.e. the enrollees in the modified incentives group) will be better motivated or prepared for weight loss than those not required to post such a bond (i.e. those in the control and standard incentives group).

In other words, we expect differential selection – those unwilling to post a forfeitable bond are excluded from the modified incentives group but are not excluded from the control group or standard incentives group. As a result, we expect that the modified incentives group will be better prepared or motivated for weight loss than the other groups.

Hypotheses Regarding Attrition

Those unsuccessful at weight loss are more likely to attrite. Participants enroll with incomplete information about certain costs and benefits of participating. Those that lose relatively little weight may update their prior beliefs and conclude that it is utility maximizing for them to drop out of the intervention.

There will be lower attrition in the program that required employees to post bonds that are refundable based on achievement of weight loss goals. Those willing to post a bond are expected to be more motivated or determined to lose weight. Selection aside, bonds may also increase retention.

Hypotheses Regarding Weight Loss

Weight loss will be greater in the standard and modified incentives groups than in the control group. Both the standard incentives group and the modified incentives group were offered financial rewards for weight loss, whereas the control group were not offered any. In other words, we hypothesize that financial rewards are effective in promoting weight loss.

In quarter 4 weight loss will be greater in the modified incentives group than in the standard incentives group.

This prediction is based on the magnitudes of the incentives; the modified incentives group has much greater incentives for 5% and 10% weight loss by the end of quarter 4. Specifically, the modified incentives group receives a refund of their \$119.40 bond if at least 5% of weight is lost, with an additional bonus of \$100 if 10% of weight is lost. Moreover, those achieving very high weight loss will be in competition for the \$250

bonus for being the “biggest loser”. In contrast, the standard incentives group is paid \$5 per month for losing 5% of starting weight and \$10 per month for losing 10% of starting weight (triple those amounts for the entire quarter). Relative to the standard incentives, the modified incentives create greater incentive for weight loss by the end of quarter four.

In addition, there are two reasons that the relative performance of the modified incentives group by the end of quarter 4 might be better than one would expect based on the magnitude of the rewards alone. First, we expect differential selection - those willing to post a bond are likely better prepared or more motivated for weight loss. Second, the research literature on loss aversion indicates that people are more motivated by a risk of losing their own money (as in the modified incentives group) than they are by the prospect of winning someone else’s money (as in the standard incentives group).

In quarters 1 through 3, weight loss will be greater in the standard incentives group than in the modified incentives group.

This prediction is also based on the magnitudes of the incentives. In quarters 1 through 3, the standard incentives group is offered \$5 per month for 5% weight loss, and \$10 per month for 10% weight loss (see Table 1 for the full schedule of financial rewards). In contrast, there is no marginal reward for 5% or 10% weight loss in any of the first three quarters for the modified incentives group (those losing any weight at all are eligible for lottery prizes, but there is no additional reward for any weight loss above the trivial amount that makes one eligible for the lottery).

However, there are three reasons that the relative performance of the modified incentives group in quarters one through three might be better than one would expect based on the magnitude of the rewards alone. The first reason is differential selection.

The second reason is loss aversion; the fear of losing one's money at year's end may motivate members of the modified incentives group to lose weight in the early quarters, even when there are no quarter-specific rewards for doing so. Third, it may take more than one quarter to achieve 5% or 10% weight loss, so in order to meet their year-end goals members of the modified incentives group may have to lose weight in earlier quarters, even though they have no financial incentives for meaningful weight loss in those quarters.

Methods and Data

To answer our research questions, we would ideally randomly assign workers into two treatment groups and one control group. The first treatment group would participate in the Company X treatment with the standard incentives and the second treatment group would participate in the Company X treatment with the modified incentives. The control group would receive all aspects of the Company X treatment except the financial rewards for weight loss. The effectiveness of each schedule of financial incentives could then be measured using a difference in differences model, which would calculate the difference between the treatment group and the control group in their change in weight before and after the intervention. (If one was interested in the impact of the daily emails and call center, it could be measured by the difference in differences of the control group and another group ("usual care") that did not receive the daily emails, call center access, or any other treatment.)

A limitation of our data is that they are not the result of a randomized controlled trial. They are opportunistic data, provided to us by Company X. As a result, we face two challenges: 1) assignment to the three treatment groups is nonrandom: the incentive

schedules were chosen by the employers; 2) the participation of employees is voluntary; there is selection by employees.

Regarding problem #1 (selection by employers into different incentive schedules), we assume that this is ignorable. In other words, we assume that employer preference for incentive structure is uncorrelated with unobserved employee characteristics that affect attrition and weight loss. Company X told us that the reason that one employer requested the modified incentives schedule (with forfeitable bonds) is because the company didn't want to pay for cash rewards. This would be more problematic if the employer requested the modified schedule because the employer thought it would be more effective for their particular employees.

Regarding problem #2 (selection by employees into participation), we consider this to be a limitation for generalizing results to the entire population but not a problem in the sense that any similar intervention is also likely to be optional, and so the findings for a set of volunteers is most relevant. All of the studies in Appendix Table 1 are all based on volunteers recruited to participate in a weight loss program, and are likewise not a random sample of the general population. To control for differential selection across incentive schedules we will control for a proxy for unobserved motivation and readiness for weight loss: the intensity of the physical activity program chosen by the enrollee at baseline.

An additional problem is that there is attrition from the program. Weight-loss interventions in general (even those without financial rewards) typically have substantial attrition (Ware 2003; Gadbury et al. 2003). There are several strategies for handling the attrition when evaluating interventions. The definitive is the intent-to-treat analysis,

which includes all patients in their groups, regardless of whether they received the treatment, deviated from the protocol or withdrew (Ware 2003). However, to implement this one must have follow-up data on all of the dropouts, which is not available in this case. Another option is to conduct a “completers” analysis, which examines data only for those who completed the study. This is likely to be biased toward showing an impact of the treatment, as those most likely to quit are probably those for whom the intervention was least effective (Ware 2003). Another option is last-observation-carried-forward, which assumes that the dropouts remained at their last measured weight. This also likely results in upward bias in estimates of program effectiveness, as weight regain is common (Ware 2003; Serdula et al. 1999). Another option is baseline-carried-forward, which assumes that after attriting the subjects return to their baseline weight. This may cause downward bias in the estimate of efficacy, as weight regain may be incomplete or slow. We present findings for both completers analysis and baseline-carried-forward.

Our data cover seven employers: five employers (with a total of eleven participating worksites) that implemented the standard set of incentives shown in Table 1, one employer (with two participating worksites) that implemented the modified incentives shown in Table 2, and one employer (with two participating worksites) that implemented the Company X intervention without any financial incentives (the control). The total number of employees in the dataset is 2,351: 1,374 facing the standard incentives, 835 facing the modified incentives, and 142 with no financial incentives. The data cover 2004-2007. We drop from the sample participants with baseline BMI below 25 because they were not eligible for financial rewards. Fourteen participants in the

control group were dropped because they were simultaneously participating in another workplace weight-loss intervention.

To examine attrition and test our hypotheses related to attrition, we estimate attrition rates by quarter and group, and also estimate hazard models of attrition as a function of age, age squared, percent of baseline weight lost at last weigh-in (if one never weighed in again after baseline this is equal to zero) and indicator variables for male, obese, morbidly obese (the reference category is overweight), group (standard incentives, modified incentives, with the control group as the reference), chose intermediate exercise program, chose advanced exercise program (reference category is chose the foundation exercise program, which is the easiest). For the hazard models of attrition, data on all four quarters are pooled.

To examine weight loss and test our hypotheses related to it, we graph the distribution of weight loss by group and quarter, both for a completers analysis (ignoring dropouts) and assuming that dropouts return to baseline (baseline carried forward). We also calculate the unconditional mean loss in pounds and percent of baseline weight lost by group and quarter, for both a completers analysis and assuming baseline carried forward. Finally, we also estimate regressions of weight loss (both pounds and percent of weight lost since baseline¹⁰) by group and quarter, for both a completers analysis and assuming baseline carried forward. The weight loss models include the following regressors: age, age squared and indicator variables for male, obese, morbidly obese (the reference category is overweight), group (standard incentives, modified incentives, with

¹⁰ When the dependent variable is a percentage that ranges between 0 and 1, researchers sometimes opt to use the logit transformation of the dependent variable before estimating OLS models. Because our dependent variable can be negative (i.e., weight can be gained), we do not use the logit transformation and instead regress the raw percentage change in weight since baseline on the regressors.

the control group as the reference), chose intermediate exercise program, chose advanced exercise program (reference category is chose the foundation exercise program), and quarter.

In all models that are estimated using a sample of all quarters pooled, standard errors are cluster-corrected at the individual level, to account for the correlation in error terms across multiple observations of the same individual.

Empirical Results

Descriptive statistics:

Table 3 presents the summary statistics for participants by group. Our overall sample (N=2,351) consists of 1,374 participants in the standard incentives group, 835 participants in the modified incentives group, and 142 participants in the control group.

In each of these groups, men are a minority: 15.9% of the standard incentives group, 21.4% of the modified incentives group, and 35.0% of the control group. The average age in each group is in the early-to-mid 40s, and average baseline BMI ranges between 31 and 33 across the groups. In each group there is a strikingly high prevalence of morbid obesity (BMI of greater than or equal to 40). In the U.S. as a whole, 4.8% of the population is morbidly obese (Ogden et al. 2006); this implies that they represent 7.3% of the 66% of Americans who are at least overweight. In contrast, the morbidly obese constitute 29.3% of the standard incentives group, 31.1% of the modified incentives group, and 21.0% of the control group.

Hypotheses Regarding Enrollment

We hypothesized that:

There will be lower enrollment in the program that required people to post forfeitable bonds. Table 4 lists the percent of the workforce that enrolled in the program, by incentive schedule. Ideally we would know the number of employees with BMI of 25 or higher, because only they are eligible for financial rewards for weight loss. Instead, for both the numerator and denominator we know only the total number of employees (i.e. those of all BMI). As a result, these are likely to be underestimates of the percentage of those eligible for financial rewards who enrolled in the program. Percent enrollment was 17.8% for the modified incentives (which required a bond), 28.4% for the standard incentives, and 20.3% for the program that offered no financial rewards for weight loss but all of the other program elements (i.e. the control group). The point estimates of enrollment are consistent with our prediction that the requirement of a bond would result in lower enrollment, but the differences are not statistically significant.

Those who are willing to post a forfeitable bond (i.e. the enrollees in the modified incentives group) will be better motivated or prepared for weight loss than those not required to post such a bond (i.e. those in the control and standard incentives group). There are two variables that can give us information about the degree of such differences in selection. The first variable is the level of exercise regimen that the employee chose at the beginning of the program. If those willing to pay the monthly fees in the modified incentives group are more motivated or prepared to lose weight, one should find that they are less likely to choose the easiest exercise regimen. This is confirmed by the data. The easiest exercise regimen (called Foundation) was chosen by 60.6% of the standard incentives group but only 55.1% of the modified incentives group (see Table 3), a difference significant at the 1% level. We also expected that the control

group, offered \$20 if they participated for the full year, would be less motivated on average and therefore more likely to choose the easiest exercise regimen than those in the modified incentives group, but we do not find this – an even lower percentage of the control group than the modified incentives group (50.0% versus 60.6%) chose the easiest exercise regimen, but the difference is not statistically significant.

The second variable that sheds light on difference in selectivity is the percentage of the program emails that the enrollees read. If those willing to pay the monthly fees in the modified incentives group are more motivated or prepared to lose weight, one should find that they read a higher percentage of the program emails. That prediction is confirmed by the data – the average percentage of emails read was 50.3% for members of the modified incentives group compared to 45.4% for members of the standard incentives group (see Table 3), a difference significant at the 1.3% level. (A caveat is that this variable is missing for 59.6% of the standard incentives group.)

The control group, being paid to participate, had the lowest email open rate of 27.97%, which is significantly different from both other groups at the 1% level. It is interesting that the control group had the lowest percentage choosing the easiest exercise regimen (which suggests more motivation or better preparation) but the lowest email open rate (which suggests lower commitment).

Overall, the patterns of both exercise regimen and email opening suggest that the group required to post a bond was selected to be better prepared and more serious about weight loss than the standard incentives group, and therefore should be less likely to attrite and more likely to lose weight.

Attrition

Table 5 lists the cumulative percentages dropping out, by quarter, for each group. In the standard incentives group, 46.4% of baseline participants have dropped out by the end of quarter 1, and cumulative attrition rises in the three subsequent quarters to 59.4%, 70.7% and 82.9%. In the modified incentives group, attrition is lower but still substantial: 28.1% after one quarter, rising in the three subsequent quarters to 33.7%, 40.0%, and 57.8%. Even in the control group, where participants are promised \$20 if they weigh in every quarter for a year, attrition is substantial: 23.2% after one quarter, rising in the three subsequent quarters to 35.9%, 41.5%, and 45.8%.

Attrition is typically substantial in weight loss interventions of all kinds (Ware 2003; Gadbury et al. 2003). However, the attrition in these groups is particularly high. For example, a recent review (Paul-Ebhohimhen and Avenuell, 2007) of RCTs involving financial rewards for weight loss found that the maximum attrition in any such study was 57.9% at 13 months, far below what the standard incentives group experienced in 12 months (82.9%) but roughly equal to what the modified incentives group experienced at 12 months (57.8%). This suggests that real-world interventions may experience far higher rates of attrition than those overseen by researchers (who for the purposes of data quality undertake extensive efforts to keep enrollees from attriting), which raises questions about how well the results of pilot studies such as those in Appendix Table 1 can be duplicated on a larger scale.

When considering the levels of attrition, one should keep in mind that enrollees were already a select sample. Participation was optional, and most employees declined to enroll.

Table 6 lists the results of multivariate models of attrition. Table cells contain relative risks of attriting, with z statistics in parentheses. The first column presents results estimated using a sample of all three groups pooled. Subsequent columns present results from models estimated separately by group. More detail on the hazard model results is provided below, where we discuss our hypotheses.

Those unsuccessful at weight loss are more likely to attrite. In the results for the pooled sample (column 1), those who have lost a higher percentage of their baseline weight are significantly less likely to attrite. Each percentage point of lagged weight loss (which is measured as the percent change between baseline and the most recent weigh-in) is associated with a 5.0% (1-.95) lower probability of attriting. This, because it suggests that the dropouts are those who are failing to lose weight, confirms the appropriateness of using the baseline carry-forward approach instead of a completers analysis to deal with attrition; those who attrite are not “missing at random” with respect to the outcome of interest.

There will be lower attrition in the program that required employees to post bonds that are refundable based on achievement of weight loss goals. Table 5 lists the unconditional attrition by group and quarter. In the modified incentives group (in which forfeitable bonds were required), dropouts totaled 28.1% in the first quarter, 33.7% in the second quarter, 40.0% in the third quarter, and 57.8% in the fourth quarter. In the standard incentives group, dropouts totaled 46.4% in the first quarter, 59.4% in the second quarter, 70.7% in the third quarter, and 82.9% in the fourth quarter. In every quarter, the amount by which attrition is lower in the group that posted a forfeitable bond is statistically significant at better than a 1% level.

Results of hazard models are presented in Table 6. The relative risk of attrition for the standard incentives group (229.9%) is significantly higher than that of the modified incentives group (111.2%), both relative to the control group (100% by construction). Thus, both the unconditional attrition rates by quarter and the hazard model results are consistent with lower attrition in the modified incentives group, who were willing to post forfeitable bonds, relative to the standard incentives group, who had no “skin in the game”. It is also consistent with our findings with regard to physical activity readiness and the email open rates, which suggested that those in the modified incentives group were selected to be better prepared and more motivated to lose weight. It is impossible to tell with these data whether the difference in attrition is due to selection or loss aversion.

Other correlates of attrition Those with higher starting BMIs are more likely to quit. In Table 6, column 1, which uses data for all three groups, the obese are 12.9% more likely to attrite, and the morbidly obese are 14.0% more likely to attrite, than the overweight. (When we replace the clinical weight classifications with a linear measure of BMI (results not shown), we find that each additional unit of BMI above the mean is associated with a 0.8% higher probability of attriting.) In addition, those who choose the advanced exercise program are 16.1% (1-.839) less likely to attrite than those who choose the easiest exercise program; this supports our earlier use of exercise program as a proxy for unobserved readiness or motivation for weight loss.

Weight Loss

Distribution of Individual Weight Loss

The distribution of percent weight loss is shown in Figures 2 (for the standard incentives group), Figures 3 (for the modified incentives group), and Figures 4 (for the control group). Percent weight loss is rounded down to the nearest percentage point.¹¹ For each group, there is a separate page devoted to the data for each quarter. On each page are two graphs: the upper graph is the distribution of weight loss in a completers analysis that ignores dropouts, and the lower graph is the distribution of weight loss in a baseline-carried-forward analysis that assumes that every dropout returned to their baseline weight. A comparison of the upper and lower graphs confirms that how attrition is handled has a substantial impact on estimated weight loss. In the upper graphs (the completers analysis), the distribution of outcomes seems more favorable (although the modal outcome is usually zero weight loss), but in the lower graphs (baseline carry-forward), by far the most common outcome is that respondents lost zero weight (largely driven by the assumption of setting dropouts at baseline weight).

Figure 3-4 indicates that more people in the modified incentives group are just over the thresholds of 5% weight loss (at which participants are refunded their year's worth of fees, or \$119.40) and 10% weight loss (at which they also receive a \$100 bonus), then just under the thresholds. This is most apparent in quarter 4, when financial rewards are tied to those thresholds for the modified incentives group, relative to earlier quarters when there are no financial rewards tied to those thresholds. Such heaping is not apparent in the distribution associated with the standard incentive schedule, which has more continuous reward thresholds.

¹¹ We round down so that everyone indicated as having a specific percent weight loss received exactly the reward associated with that percent weight loss. If we rounded to the nearest percentage point, a participant who lost 4.6% of her starting weight would be rounded to 5% even though she would not have qualified for the financial reward associated with achieving 5% weight loss.

We next discuss the evidence regarding our hypotheses regarding weight loss.

Weight loss will be greater in the standard and modified incentives groups than in the control group. We first test for differences in unconditional means of weight loss in pounds and percent of baseline weight by quarter and group, then present regression results of weight loss (pounds or percent of baseline weight) on group, controlling for observable characteristics. We then test for differences in unconditional probability of losing 5% and 10% of baseline body weight, then present results of probit regressions in which the dependent variable is an indicator for losing 5% or 10% of baseline weight, controlling for observable characteristics. Note that the differences between the treatment groups and the control group can be interpreted as the effect of the financial incentives, distinct from all of the other program elements (e.g. daily emails and call center support) shared by the control group.

Table 7 lists weight loss in pounds and in percent of baseline weight, by group and quarter. The cells also list the minimum and maximum weight loss (a negative minimum weight loss implies weight gain) for that group in that quarter (the minimum and maximum are not affected by how dropouts are treated, so they are entered in only the leftmost column for each group). Because so many participants attrite, and attrition is correlated with weight loss success, estimates of average weight loss are extremely sensitive to how attrition is handled. We focus here on the baseline carried forward analysis, which assumes that everyone who dropped out went back to their baseline weight.

Average weight loss in the control group is 2.8 pounds (1.4%) in the first quarter, 2.2 pounds (1.1%) in the second quarter, 2.1 pounds (1.1%) in the third quarter, and 2.1

pounds (1.1%) in the fourth quarter. Relative to the treatment group, average weight loss is significantly higher in the modified incentives group: 3.0 pounds (1.5%) in the first quarter, 3.2 pounds (1.5%) in the second quarter, 2.5 pounds (1.2%) in the third quarter, and 3.6 pounds (1.8%) in the fourth quarter. However, we fail to reject the null hypothesis of no difference between the control and standard incentives group; in fact, mean estimates are consistently higher for the control group than the standard incentives group.

Despite the small average weight loss in the standard incentives group, there are some substantial success stories; the maximum weight lost since baseline is 56.4 lbs. in quarter 1, 89.4 lbs. in quarter 2, 109.2 lbs. in quarter 3, and 101.4 lbs. in quarter 4. For any given mean, success stories are balanced by failures; for example, the maximum weight *gain* is 12.6 lbs. in quarter 1, 25 lbs. in quarter 2, 56 lbs. in quarter 3, and 18.2 lbs. in quarter 4.

We present in Tables 8A (completers analysis) and 8B (baseline carried-forward), results of regression models in which the dependent variable is pounds of weight loss, and in Tables 9A (completers analysis) and 9B (baseline carried-forward) we present the results of models in which the dependent variable is percent of baseline weight lost.

In the baseline carried forward analysis, the modified incentives group lost 1.7 pounds more than the control group (Table 8B) and 0.8% more of baseline weight than the control group (Table 9B) by the end of quarter 4. Contrary to our prediction, the standard incentives group did not lose significantly more weight than the control group by the end of quarter 4; again, the point estimate suggests that they actually gained weight relative to the control group, although this change is not statistically significant.

(If one ignores dropouts (Tables 8A and 9A), the standard incentives group lost weight relative to the control group, but this is not statistically significant.)

However, while there is a significant effect found at the end of quarter four, there is no statistically significant difference in weight loss (either pounds or percent) between the control group and either the standard or modified incentives group in quarters one, two, or three, or when all four quarters are pooled.

We also measure weight loss by success in reaching certain benchmarks. Table 10 lists the percent of participants losing 5% of baseline weight, by group and quarter. In the baseline carried forward analysis, the percentage of the control group that lost 5% of their baseline weight, by quarter, was: 9.2%, 8.4%, 15.5%, and 12.7%. This is greater than the corresponding percentages for the standard incentives group (7.6%, 7.9%, 7.3%, and 4.1%) but less than those for the modified incentives group (11.9%, 15.8%, 13.8%, 19.3%). Surprisingly, in each quarter, members of the control group had a higher probability of losing 5% of baseline weight than members of the standard incentives group; in quarter two this difference is significant at the 10% level.

Tables 11A and 11B present the results of probit regressions in which the dependent variable is an indicator for whether the respondent achieved 5% weight loss. As before, the “A” table presents results of a completers analysis while the “B” table presents results of a baseline-carried forward analysis. In the baseline carried forward analysis of Table 11B, those in the modified incentives group are 5.3 percentage points more likely than the control group to lose 5% of baseline weight by the end of quarter four (which is statistically significant at the 5% level), but those in the standard incentives group are 8.5 percentage points less likely than the control group to lose 5% of

baseline weight by the end of quarter 4 (which is significant at the 1% level). Differences between the control group and treatment groups are weaker in quarters one, two, and three, and in the model estimated with all four quarters pooled.

We also examine the probabilities of losing 10% of baseline weight, the outcome that the NIH (1990) recommends for evaluating weight loss programs. Table 12 lists the unconditional probabilities of losing 10% of baseline weight by group and quarter, for both completers and baseline carried forward analyses. Assuming dropouts returned to their baseline weight, the percentage of the control group that lost 10% of baseline weight, by quarter, was 0.7%, 2.8%, 3.5%, and 4.2%. These are generally higher than the corresponding percentages for the standard incentives group (1.0%, 2.0%, 2.6%, and 1.7%) and comparable to those for the modified incentives group (1.9%, 4.3%, 4.1%, and 6.9%).

Results of probit regressions in which the dependent variable is an indicator for 10% weight loss are presented in 13A and 13B (10% weight loss). As before, tables labeled A present results of completers analysis while those labeled B present results from baseline carried forward analysis.

Results of probit regressions in which the dependent variable is an indicator for 10% weight loss are presented in 13A and 13B (10% weight loss). As before, tables labeled A present results of completers analysis while those labeled B present results from baseline carried forward analysis. In general, there are not significant differences in the probability of such weight loss between the control group and either treatment group; the exception is that in the completers analysis, those in the modified incentives group are

9 percentage points more likely than the control group to lose 10% of baseline weight by the end of quarter 4.

In quarter 4 weight loss will be greater in the modified incentives group than in the standard incentives group. This is true for the unconditional means in Table 7. Average weight loss is 3.6 pounds (1.8%) in the modified incentives group compared to 1.0 pounds (0.5%) in the standard incentives group, a difference that is statistically significant at the 1% level. The regression analysis reveals that, by the end of quarter 4, those in the modified incentives group lost an additional 1.7 pounds (Table 8B) or 0.8% of baseline weight (Table 9B) relative to the control group.

Table 10 indicates that at the end of quarter 4, 19.3% of the modified incentives group had lost 5% or more of their baseline weight, compared to only 4.1% of the standard incentives group, a difference significant at the 1% level. In the probit regressions for 5% weight loss assuming baseline carried forward (Table 11B), those in the modified incentives group were 13.8 percentage points more likely than those in the standard incentives group to lose 5% of baseline weight.

Table 12 shows that the percent losing 10% or more of baseline weight was 6.9% in the modified incentives group and only 1.7% in the standard incentives group, a difference significant at the 1% level. In the probit regressions for 10% weight loss (Table 13B), those in the modified incentives group were 4.7 percentage points more likely than those in the standard incentives group to lose 10% of their baseline weight by quarter 4, a difference that is statistically significant.

In quarters 1 through 3, weight loss will be greater in the standard incentives group than in the modified incentives group.

Contrary to our prediction, weight loss is greater in the modified incentives group than in the standard incentives group in quarters one through three. Table 7 shows that those in the modified incentives group lost an average of 3.0, 3.2, and 2.5 pounds in the first three quarters, compared to the standard incentives group average losses of 1.9, 1.7, and 1.7 pounds. This difference is statistically significant at the 1% level in quarters one and two and is significant at the 5% level in quarter three.

This is also reflected in the regression results. Contrary to our prediction, those in the modified incentives group consistently lose more pounds (Table 8B) and a greater percentage of baseline weight (Table 9B) relative to those in the standard incentives group, but this is less evident when one ignores dropouts (Tables 8A and 9A).

Table 10 shows that in each case a higher proportion of the modified incentives group than the standard incentives group achieved 5% weight loss: 11.9% versus 7.6% in quarter one, 15.8% versus 7.9% in quarter two, and 13.8% versus 7.3% in quarter three; in each case these differences are statistically significant at the 1% level. In the probit regressions for 5% weight loss assuming baseline carried forward (Table 11B), those in the modified incentives group were more likely than those in the standard incentives group to lose 5% of baseline weight in quarters one, two, and three.

Table 12 shows that the probability of losing 10% or more of baseline weight is consistently higher in the modified incentives group than the standard incentives group in quarters one, two, and three. These differences are significant at a level of 5% or better. In the probit regression of Table 13B, the modified incentives group consistently from quarters one through three has a higher probability than the standard incentives group of losing 10% of baseline weight; this difference is statistically significant in quarter 2.

In summary, the evidence contradicts our hypothesis that the standard incentives group should lose more weight than the modified incentives group in quarters one, two, and three. This suggests that the effect of greater financial incentives for the standard incentives group is swamped by some combination of more favorable selection into the modified incentives group, loss aversion, and the necessity of starting early to achieve 5% or 10% weight loss by the end of quarter four.

Effect of the non-financial program elements

The effect of the non-incentive aspects of the Company X treatment (emails, call center access, and weight-ins) can be estimated by comparing the weight loss in the control group to the changes in weight experienced by those eligible for weight-loss programs but receive no treatment. We have none of the latter in our data, but Jeffery, Wing, et al. (1993), in their evaluation of an intensive behavior therapy program with incentives, had a control group that received no treatment of any kind; it experienced virtually no change in average weight after 6 or 12 months. To the extent those findings generalize, it suggests that the changes in weight observed in the control group are entirely attributable to the Company X treatment. Assuming baseline carried forward, the average changes in weight in the control group, by quarter, are: 2.8 pounds (1.4%), 2.2 pounds (1.1%), 2.1 pounds (1.1%) and 2.1 pounds (1.1%); see Table 7.

Other correlates of weight loss

A consistent finding across all of the weight loss regressions reported in Tables 8, 9, 11, and 13 is that, all else equal, men tend to lose more weight than women. In the baseline carried forward analyses, we find that at the end of quarter four men have lost an additional 1.7 pounds (Table 8B) or 0.6% of baseline weight (Table 9B), and have a 4.0

percentage point higher probability of losing 5% of baseline weight (Table 11B) and a 2.7 percentage point probability of losing 10% of baseline weight (Table 13B). The previous literature on correlates of weight loss has also tended to find that men lose more weight than women (e.g. Williamson et al. 1992; Stunkard and McLaren-Hume 1959).

Conclusion/Discussion:

A 2006 Institute of Medicine report on preventing obesity set the immediate next step – which it described as an essential priority action for the near future – as “learning what works and what does not work and broadly sharing that information.” (IOM 2006, p. 410). It also notes that “All types of evaluation can make an important contribution to the evidence base upon which to design policies, programs, and interventions.” (Ibid, p. 4). This paper makes a contribution to that effort by documenting attrition and weight loss in a large program that offers financial incentives for weight loss.

The program studied is of particular interest because it is a real-world intervention, not a pilot program designed and monitored by researchers. As a result, the data are informative about how such interventions work in the real-world. However, because it is a real-world intervention, it suffers the limitations of selection by employers of incentive schedule, and a relatively small control group (142 out of a total sample of 2,351).

We study the two outcomes recommended by the NIH for evaluating weight loss interventions: attrition and weight loss. We find higher attrition (up to 82.9% after one year) than virtually all previous studies (see Appendix Table 1 and Ebhohimhen and

Avenell, 2007). Another recent study of real-world wagers on own weight loss also found 80% failure (Burger and Lynham, 2008).

We find that the financial rewards in this program are associated with modest changes in weight. After one year, those in the modified incentives group lose 1.7 pounds more than those in the control group, while the weight loss of those in the standard incentives group is not statistically distinguishable from that of the control group. The NIH considers a loss of 10% of baseline weight in 6 months to one year to be good progress for an obese individual (NHLBI 2000). By this standard, very few participants in this program achieve good progress toward weight loss: just 1.7% of the standard incentives group and 6.9% of the modified incentives group lost 10% of their starting weight in 12 months. Results of a probit regression for 10% weight loss indicate that neither the standard incentives group nor the modified incentives group is significantly more likely than the control group to lose 10% of baseline weight. By most measures, participants in the modified incentives group had 12-month weight loss that was greater than those in the standard incentives group, but it is not clear how much of this is due to selection and how much is due to bonds, controlling for selection.

The weight loss associated with the program we examine is generally smaller than that documented in the previous literature. For example, Volpp et al. (2008) estimate mean 16-week weight loss to be 13.1 lbs. when rewards take the form of a lottery with a daily expected value of \$3, and 14.0 lbs. when the rewards take the form of deposit contracts or bonds, whose amount is chosen by the enrollee but can vary between \$0 and \$3 per day and is matched 1:1 if the weight loss goal is achieved.

Our findings are closer to those of Finkelstein et al. (2007), who find modest weight loss (between 2.0 and 4.7 lbs.) at three months, but no significant weight loss at six months, associated with financial rewards that varied between \$7 and \$14 per percentage point of weight lost after six months. Likewise, Butsch et al. (2007) find no significant difference in 12-week weight loss between a treatment group offered a \$150 refund of their enrollment fee if they lost 6% of their initial weight, and a control group which was not eligible for such a refund.

Overall, our findings regarding attrition and weight loss suggest that the larger effect sizes found in pilot programs (such as those described in Appendix Table 1) may be overoptimistic about what can be achieved on a larger scale.

To put our findings in a the broader context of what works to promote weight loss, a literature review (Douketis et al. 2005) found that dietary and lifestyle therapy tends to result in less than 5 kg weight loss after 2-4 years, pharmacologic therapy results in 5-10 kg weight loss after 1-2 years, and surgical therapy results in 25-75 kg weight loss after 2-4 years. At this point, financial rewards remain a promising method for weight loss but there is not yet the scientific evidence to suggest that they can be as effective as traditional medical approaches.

To some extent, this may be because the right design of such incentives has not yet been developed. The findings of this paper, combined with those of the previous literature and research in psychology and behavioral economics, offer guidance on how to improve the effectiveness of programs that offer financial incentives for weight loss. The first suggestion is based on the old adage “you get what you pay for.” In this program, and in all the programs listed in Appendix Table 1, payoffs were based on

participants losing a specific percentage of their baseline weight. However, what employers and insurance companies really want is for the participants to be healthier – lose fat and perhaps add muscle – and therefore have lower health care costs. However, the metrics of pounds, kilograms, and body mass index do not distinguish fat from muscle (Burkhauser and Cawley 2008). Rewards for losses from starting weight are disincentives to engage in vigorous exercise which could lead to muscle gain. Future interventions should carefully consider what outcome they want to reward; in particular, the goals of the program may be better served by rewarding the loss of fat.¹²

Given that most of the participants in each treatment group dropped out, a high priority is to decrease attrition. One strategy is to pay rewards or put bonds at risk of forfeiture more often, to make them more salient and more immediately reinforcing (Ainslie 1975; Coates et al. 1982). A meta-analysis of contingency management studies found that vouchers were more effective at increasing drug abstinence when they were paid immediately upon negative test results (Lussier et al. 2006). A second strategy to reduce attrition is to pay people to participate (weigh in); through three quarters, the control group that was offered \$20 if they weighed in for all four quarters had lower attrition than the standard incentives group that was offered no such reward. A third strategy is to target those most likely to drop out for extra encouragement not to quit. This paper finds that women and those with high baseline BMI are more likely to quit. Previous research (Texiera et al. 2004) has found that dieting history, quality of life, and other factors predict attrition in weight loss programs. Such information could be

¹² One could also consider rewarding behavior change rather than weight. The advantage is that there is weak evidence that financial rewards are more effective in changing behavior than weight, but the disadvantage is that it is more difficult to monitor behavior than to monitor weight (Paul-Ebhohimhen and Avenell, 2007).

collected at baseline and individuals at risk of attriting could be targeted for special encouragement.

Such programs should also consider including a lottery component in all incentive schedules. Because individuals tend to overweight the probability of unlikely events (Kahneman and Tversky 1979) a lottery could be cheaper than guaranteed payments of the same expected value.

Bonds that are forfeited for failure to achieve goals regarding healthy behavior can be effective (Kane et al. 2004) but take-up is low (Jeffery 1978). One option is for programs to *allow* (but not require) such bonds or deposit contracts. To create an extra incentive for adherence, any forfeited moneys should go to an organization that the participant abhors. For example, stickK.com allows participants to select an “anti-charity” that would receive their forfeited bond; examples on their website include: Americans United for Life, NARAL Pro-Choice America Foundation, and the George H.W. Bush and William Jefferson Clinton Presidential Libraries.

When facilitating the posting of a bond, it may not be advantageous to have it take the form of monthly fees; this creates incentive for people to quit quickly if they feel they aren't making progress. Moreover, it requires participants to repeatedly exercise willpower to submit the next payment. Instead, a single large up-front payment should be encouraged, because that requires willpower only at one point in time, and people may welcome the opportunity to make themselves feel obligated to finish the program.

In general, there should be greater consideration of the magnitude of the financial rewards. Paul-Ebhohimhen and Avenell (2007) review the literature and conclude that “The choice of the amount, frequency or method of administration of the financial

incentives was not justified in any study.” (p. 365). Randomized studies could be conducted to determine the elasticity of weight loss to financial rewards, by type of reward (certain payment, lottery, bond).

Offering larger incentives may increase weight loss and decrease attrition for multiple reasons. Loewenstein and Prelec (1992), summarizing empirical studies of time preference, note that large dollar amounts suffer less proportional discounting than do small dollar amounts. In other words, if the year-end payoff is small, people will succumb to temptation for only a small fraction of the year-end payoff, but if the year-end payoff is large in absolute terms, a higher percentage will be required to make them succumb to temptation today. Thus, not only does a bigger payoff represent a stronger incentive, but it is discounted less, further increasing its effectiveness.

Ainslie (1975) finds evidence that a combination of small early rewards and larger late rewards can be especially effective. This suggests that a combination of token payments at each weigh-in, combined with a year-end bond refund or other type of large bonus may especially increase adherence.

A final suggestion for the design of such programs is to beware of unintended consequences. Company X told us that their client using the modified incentives schedule realized that workers were using unhealthy methods to achieve weight loss just before the end of quarter 4 because the rewards were so great. The client decided to cease using such a back-loaded incentive scheme in order to avoid such unintended consequences in the future. There is a long history of such interventions creating incentives for unhealthy behaviors. In the very first study to use deposit contracts to encourage weight loss, Mann (1972) discovered that his subjects used unhealthy weight

loss strategies to get back their forfeitable bonds: “Unsolicited anecdotal reports from some of the subjects indicated that they had used extreme measures at various times to lose weight rapidly and temporarily in order to avoid aversive consequences. These measures, reportedly, included taking laxatives, diuretics, and doing vigorous exercises just before being weighed. This problem may have occurred because the contract specified that the treatment contingencies be delivered contingent upon specified weight changes rather than the behaviors that can produce these changes (p. 108-109)”

One way to avoid creating incentives for unhealthy behaviors is to weigh participants at random times.¹³ Instead, financial rewards could be based on random weigh-ins, just as professional sports leagues rely on random drug testing.

We conclude with two directions for future research. One is to determine whether such financial incentives lose their effectiveness over time. Loewenstein et al. (2001) argue that individuals’ response to risk is strongly related to newness; people overreact to new risks and tend to underreact to familiar ones. This suggests that an incentive program may be most effective when it is first implemented, but as it becomes routine its motivational power may deteriorate. This is an empirical question that should be a high priority for future studies.

Another important area for future research is to determine whether the introduction of external incentives leads to a depreciation of intrinsic motivation. Intrinsic motivation may be costly to maintain, and individuals may rationally invest less in it when an extrinsic set of incentives is introduced. Decreased intrinsic motivation may offset some or all of the benefit of the extrinsic incentives, and when the extrinsic

¹³ Another reason for randomizing the time of weigh-ins is that if participants know they have the option to purge before a scheduled weigh-in, it may give them an excuse to not diet and exercise.

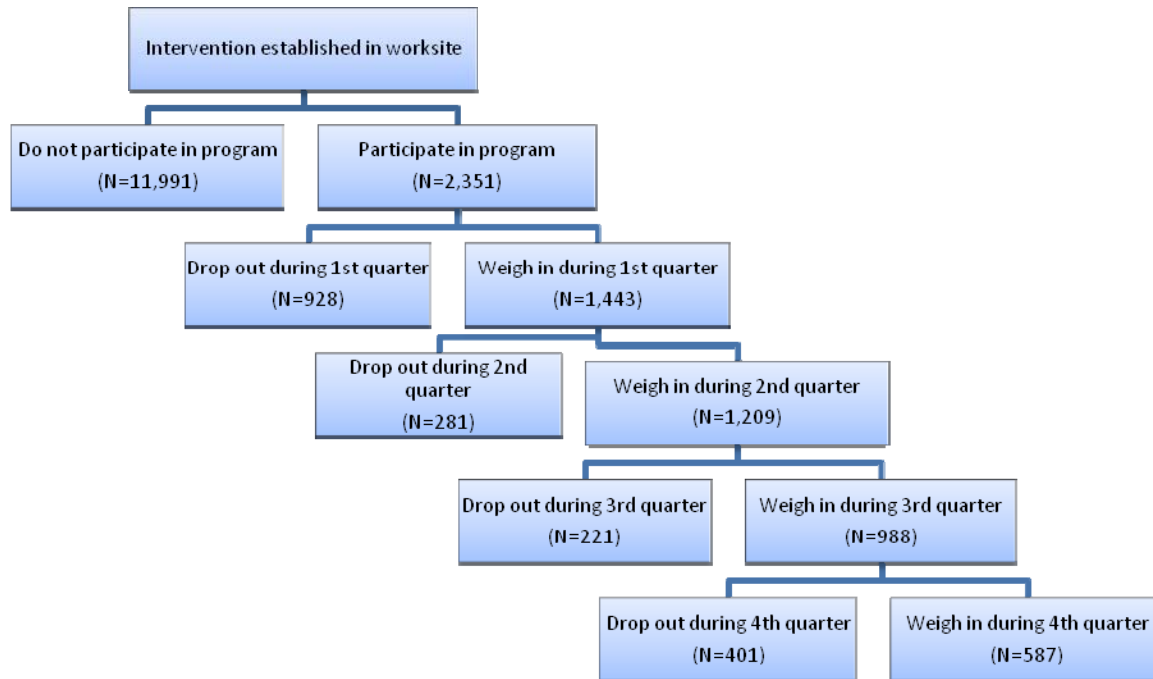
incentives are removed the participant might engage in even more myopic unhealthy behavior than before the extrinsic incentives were introduced, until intrinsic motivation is restored to its previous level (Fishbach and Trope 2005; Lepper et al. 1973; Deci 1971). In the current context, this raises the question of whether, at the conclusion of the worksite intervention when financial incentives are ended, former participants are less motivated to control their weight than before the intervention.

Works Cited (Incomplete)

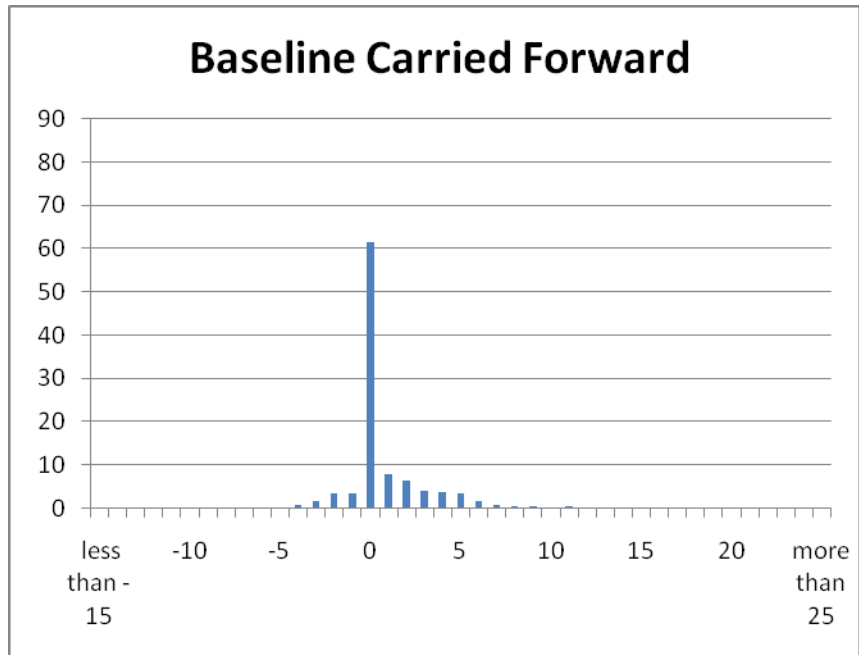
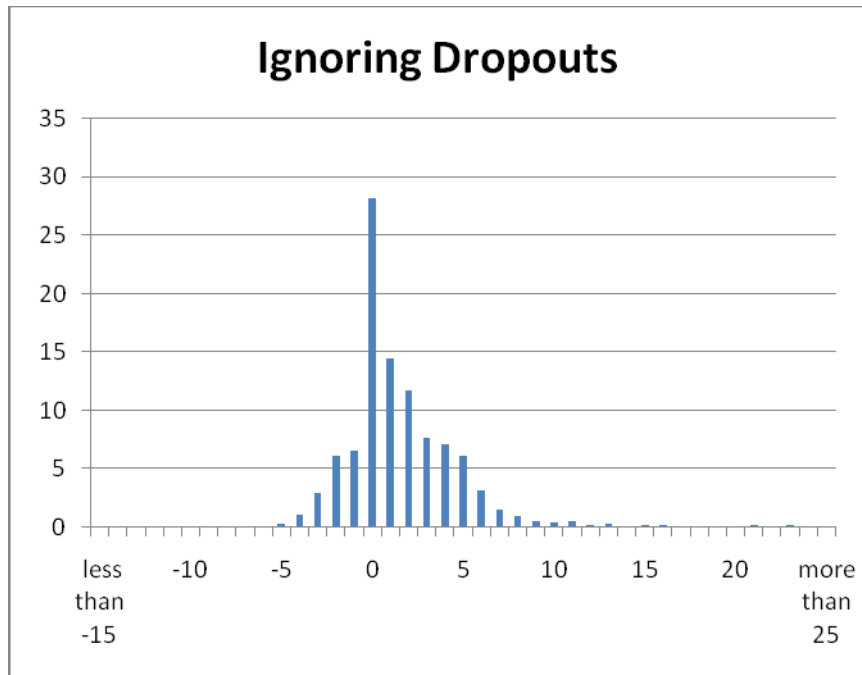
- Ainslie G. 1975. Specious reward: a behavioral theory of impulsiveness and impulse control. *Psychological Bulletin*. 82(4): 463-496.
- Bastiat F. 1850. That Which is Seen, and That Which is Not Seen.
- Beunckens, C. 2005. Direct likelihood analysis versus simple forms of imputation for missing data in randomized clinical trials. *Clinical Trials*. 2(5): 379-386.
- Bish CL, Blanck HM, Serdula MK, et al. 2005. Diet and Physical Activity Behaviors Among Americans Trying to Lose Weight: 2000 Behavioral Risk Factor Surveillance System. *Obesity Research*, 13(3): 596-607.
- Cawley, John. "The Impact of Obesity on Wages." *Journal of Human Resources*, Spring 2004, 39(2): 451-474.
- Cawley, John, John A. Rizzo, and Kara Haas. "Occupation-Specific Absenteeism Costs Associated with Obesity and Morbid Obesity." *Journal of Occupational and Environmental Medicine*, December 2007, 49(12): 1317-1324.
- Deci EL. 1971. Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*. 18(1): 105-115.
- Finkelstein EA, Fiebelkorn IC, and Wang G. 2003. National Medical Spending Attributable to Overweight and Obesity: How Much, and Who's Paying? *Health Affairs Web Exclusive* W3-219.
- Fishbach A, Trope Y. 2005. The substitutability of external control and self-control. *Journal of Experimental Social Psychology*, 41(3): 256-270.
- Ford, Earl S., David G. Moriarty, Matthew M. Zack, Ali H. Mokdad, and Daniel P. Chapman. 2001. "Self-Reported Body Mass Index and Health-Related Quality of Life: Findings from the Behavioral Risk Factor Surveillance System." *Obesity Research*, 9(1): 21-31.
- Higgins, Stephen T., Sheila M. Alessi, and Robert L. Dantona. 2002. "Voucher-Based Incentives: A Substance Abuse Treatment Innovation." *Addictive Behaviors*, 27: 887-910.
- Institute of Medicine. 2006. *Progress in Preventing Childhood Obesity: How do we Measure Up?* (National Academies Press: Washington, D.C.).
- Jeffery et al. "Use of Personal Trainers and Financial Incentives to Increase Exercise in a Behavioral Weight-Loss Program" *Journal of Consulting and Clinical Psychology*, 66(5): 777-783 (1998).
- Jeffery et al. 1993. "Strengthening Behavioral Interventions for Weight Loss: A Randomized Trial of Food Provision and Monetary Incentives." *Journal of Consulting and Clinical Psychology*, 61: 1038-1045.
- Lepper MR, Greene D, Nisbett RE. 1973. *Journal of Personality and Social Psychology*. 28(1): 129-137.
- Kahneman D, Tversky A. 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47(2): 263-292.
- Kramer et al. 1986. "Maintenance of Successful Weight Loss Over 1 Year: Effects of Financial Contracts for Weight Maintenance or Participation in Skills Training." *Behavioral Therapy*, 17: 295-301.

- Loewenstein G, Weber EU, Hsee CK, et al. 2001. Risk as feelings. *Psychological Bulletin*, 127: 267-286.
- Lussier JP, Heil SH, Mongeon JA, et al. 2006. A meta-analysis of voucher-based reinforcement therapy for substance use disorders. *Addiction*, 101: 192-203.
- NIH Technology Assessment Conference Panel. Methods for Voluntary Weight Loss and Control. *Annals of Internal Medicine*, 119(7, Part 2): 764-770.
- Ogden CL, Carroll MD, Curtin LR, et al., "Prevalence of Overweight and Obesity in the United States, 1999-2004, *JAMA*, 295 (2006): 1549-1555.
- Paul-Ebhohimhen, V. and A. Avenell. 2007. "Systematic Review of the Use of Financial Incentives in Treatments for Obesity and Overweight." *Obesity Reviews*, 9: 355-367.
- Petty, N.M., and B. Martin. 2002. "Low-Cost Contingency Management for Treating Cocaine- and Opioid-Abusing Metadone Patients." *Journal of Consulting and Clinical Psychology*, 70(2): 398-405.
- Sindelar, Jody L. 2008. "Paying for performance: the power of incentives over habits." *Health Economics*, 17: 449-451.
- Stunkard, A., & McLaren-Hume, M. (1959). The results of treatment for obesity. *Archives of Internal Medicine*, 103, 79-85.
- Tversky A, Kahneman D. 1991. Loss aversion in riskless choice: a reference-dependent model. *Quarterly Journal of Economics* 106: 1039-61.
- Vidal J. 2002. Updated Review on the Benefits of Weight Loss. *International Journal of Obesity*, 26(Suppl 4): S25-S28.
- Ware, James H. 2003. "Interpreting Incomplete Data in Studies of Diet and Weight Loss." *New England Journal of Medicine*, 348(21): 2136-2137.

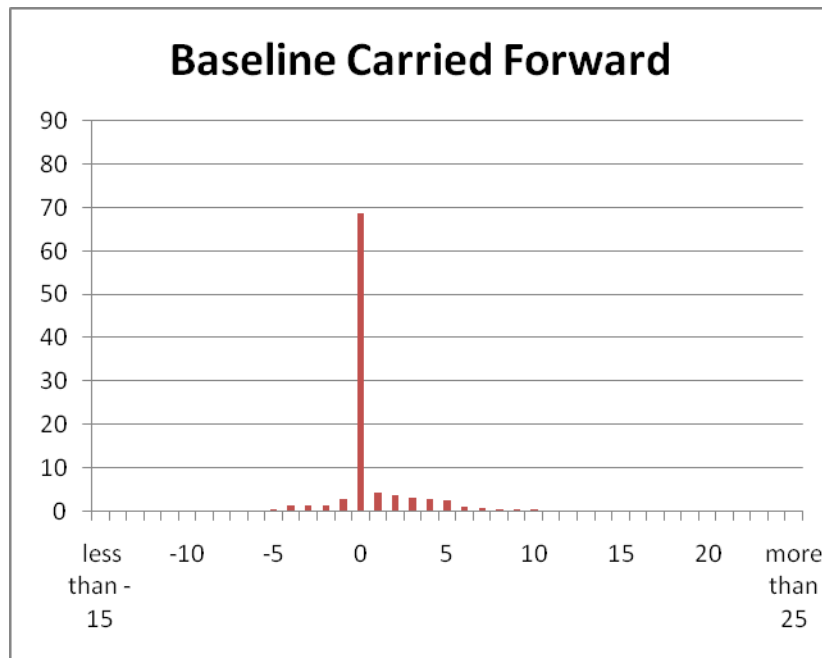
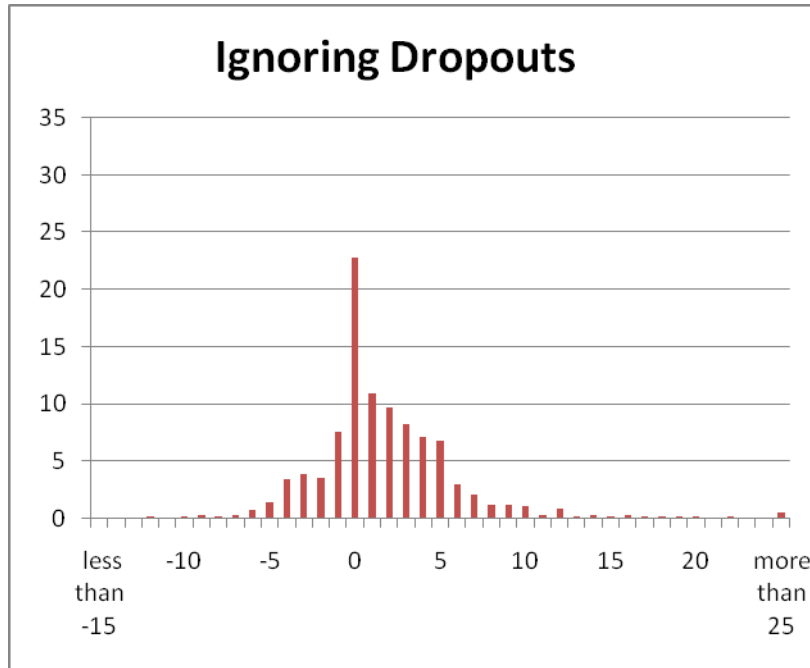
Figure 1:
Flow Diagram of Attrition and Analysis



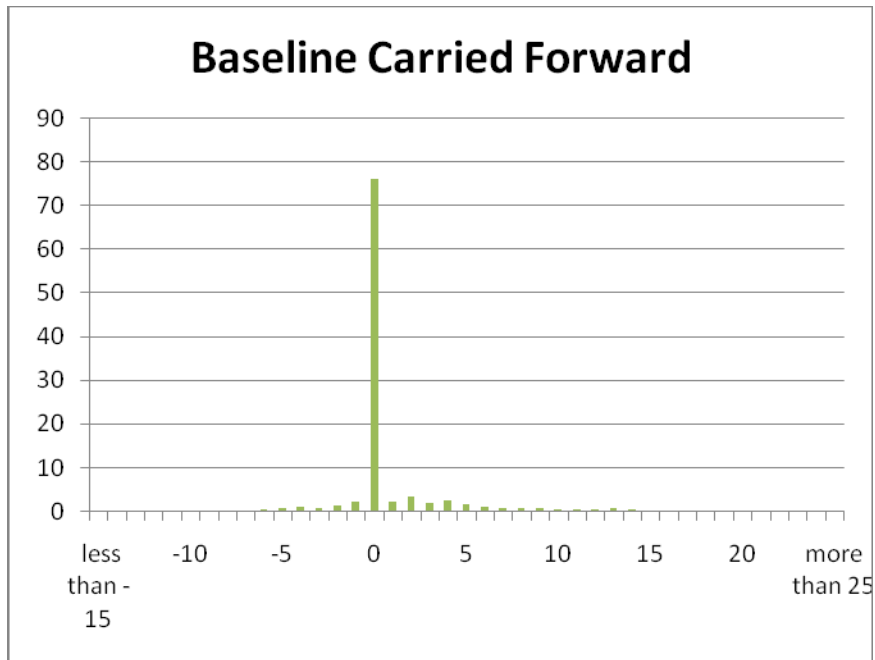
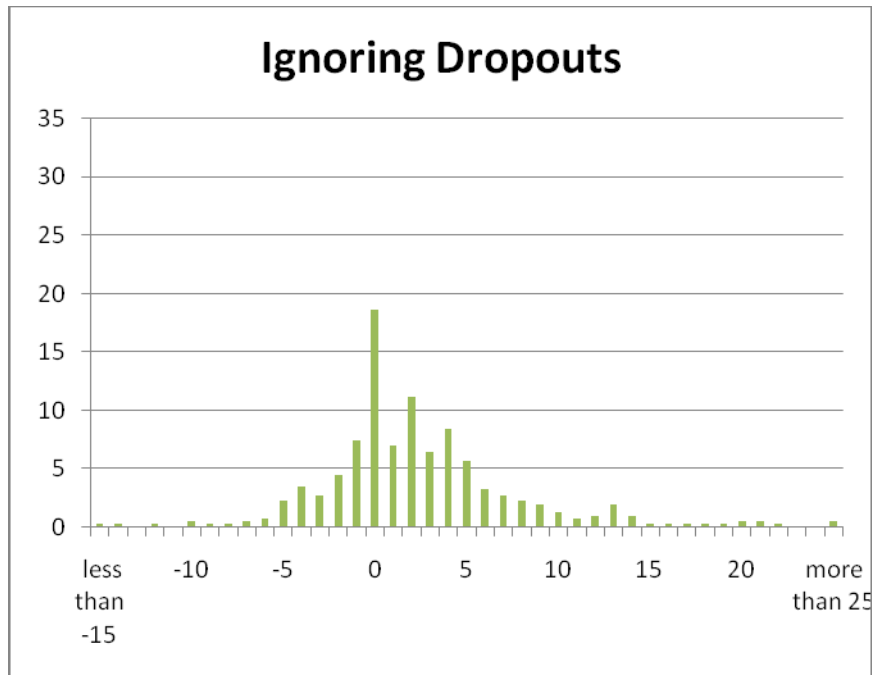
**Figure 2-1: Distribution of Percent Weight Loss under Standard incentives
Quarter 1**



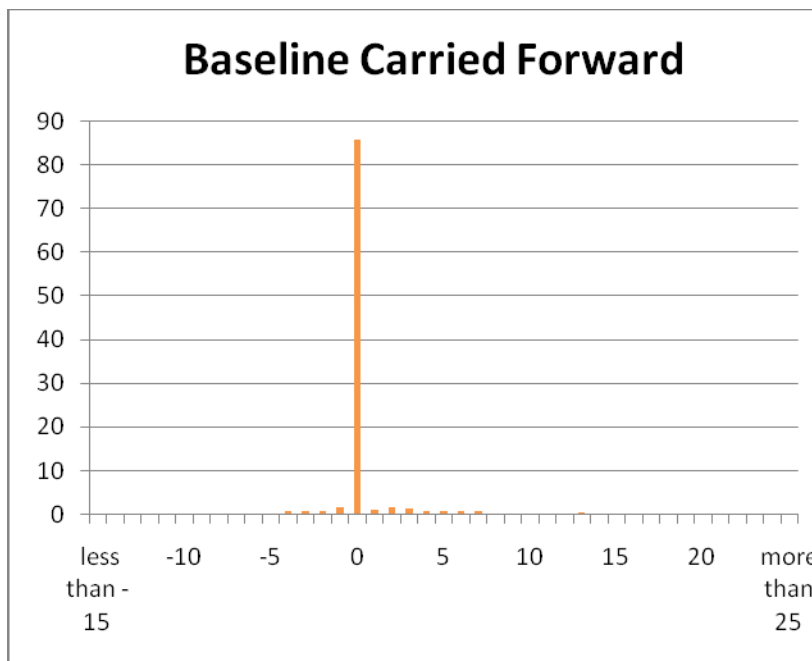
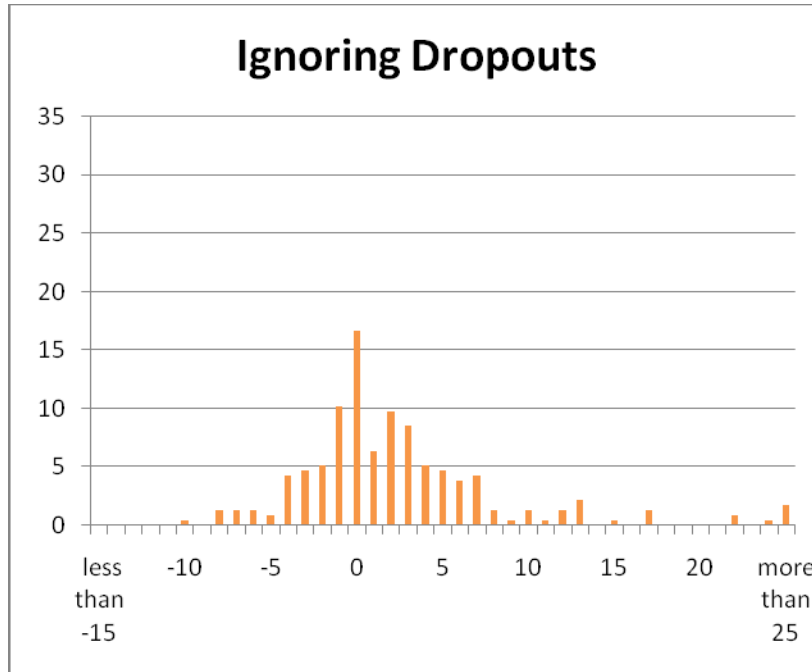
**Figure 2-2: Distribution of Percent Weight Loss under Standard incentives
Quarter 2**



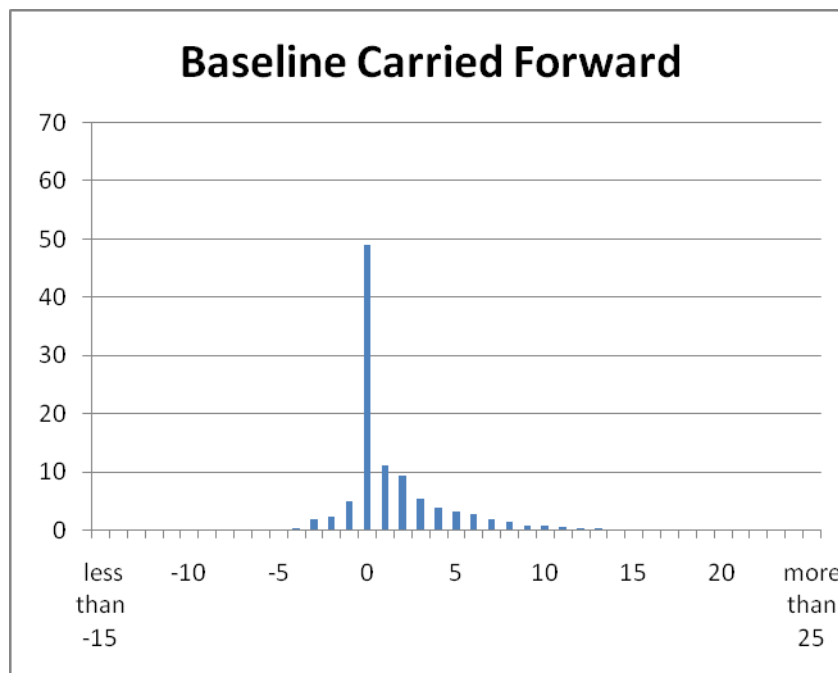
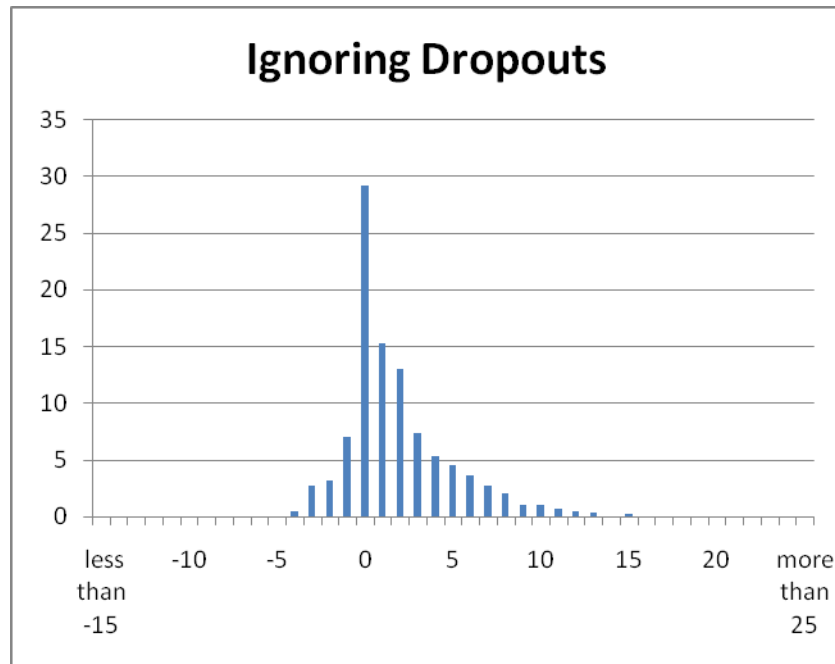
**Figure 2-3: Distribution of Percent Weight Loss under Standard incentives
Quarter 3**



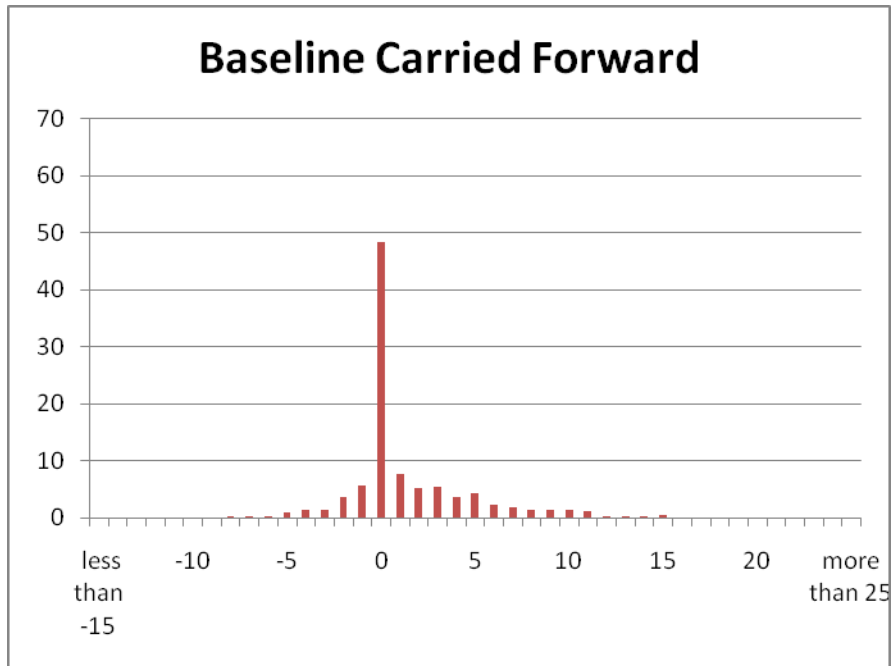
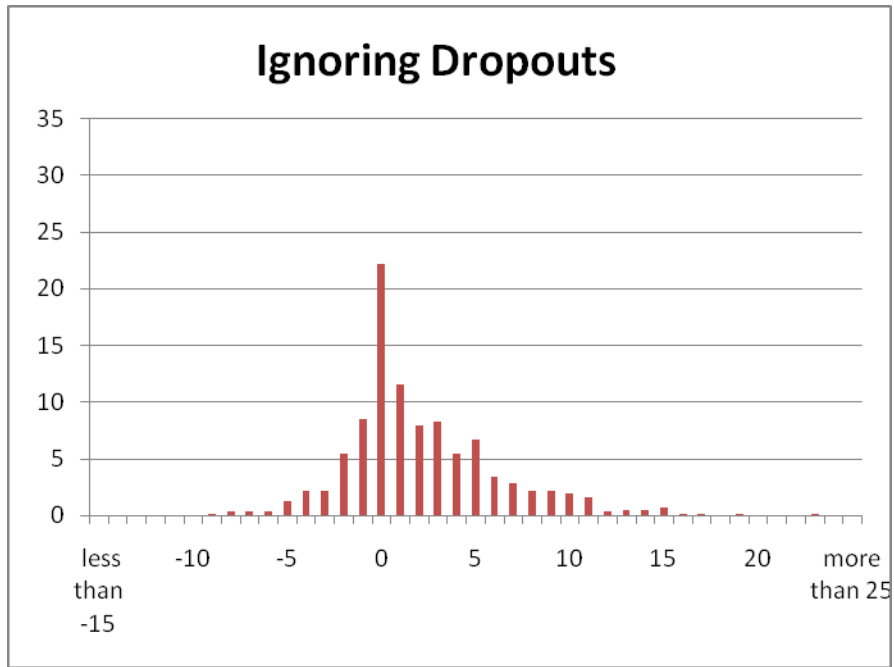
**Figure 2-4: Distribution of Percent Weight Loss under Standard incentives
Quarter 4**



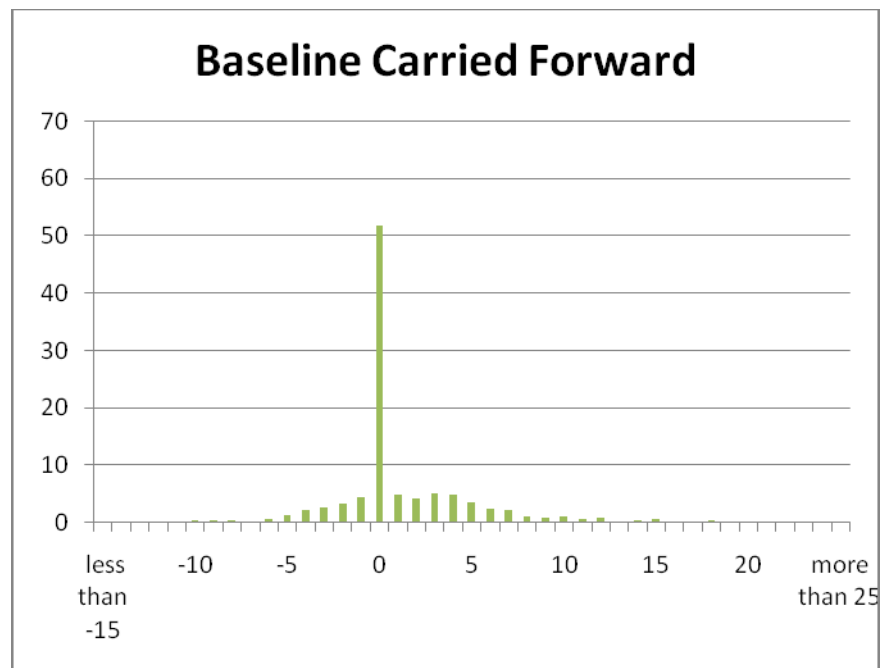
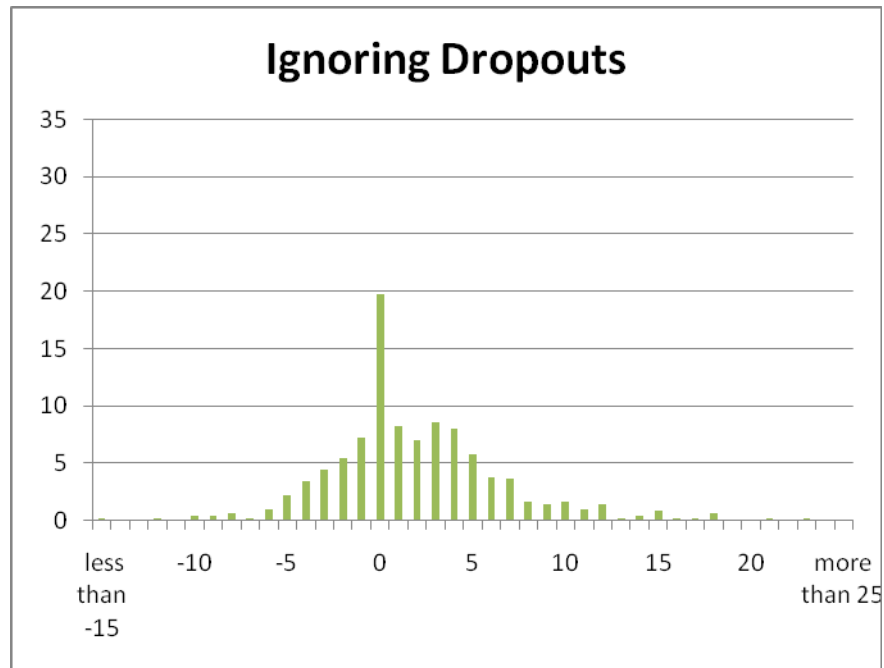
**Figure 3-1: Distribution of Percent Weight Loss under Modified incentives
Quarter 1**



**Figure 3-2: Distribution of Percent Weight Loss under Modified incentives
Quarter 2**



**Figure 3-3: Distribution of Percent Weight Loss under Modified incentives
Quarter 3**



**Figure 3-4: Distribution of Percent Weight Loss under Modified incentives
Quarter 4**

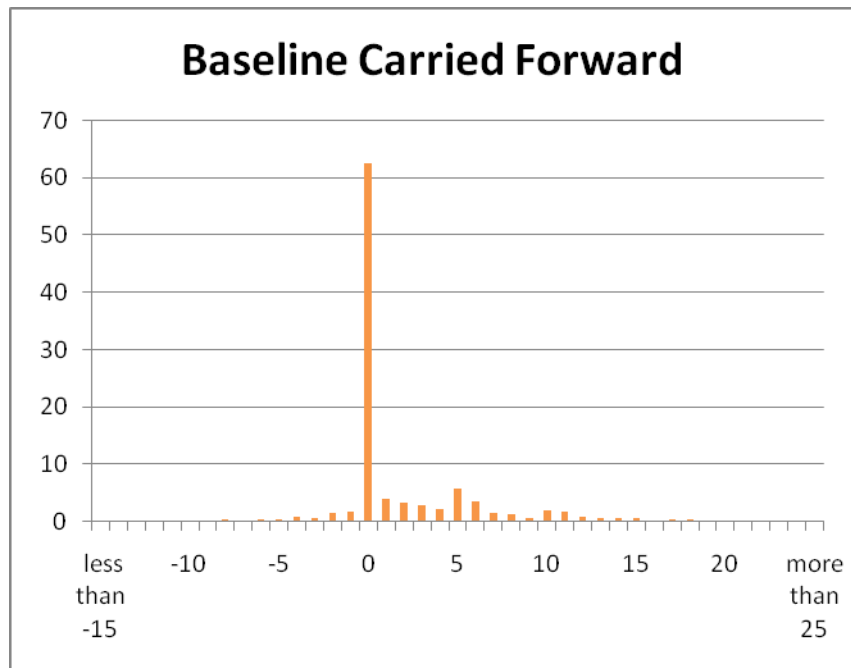
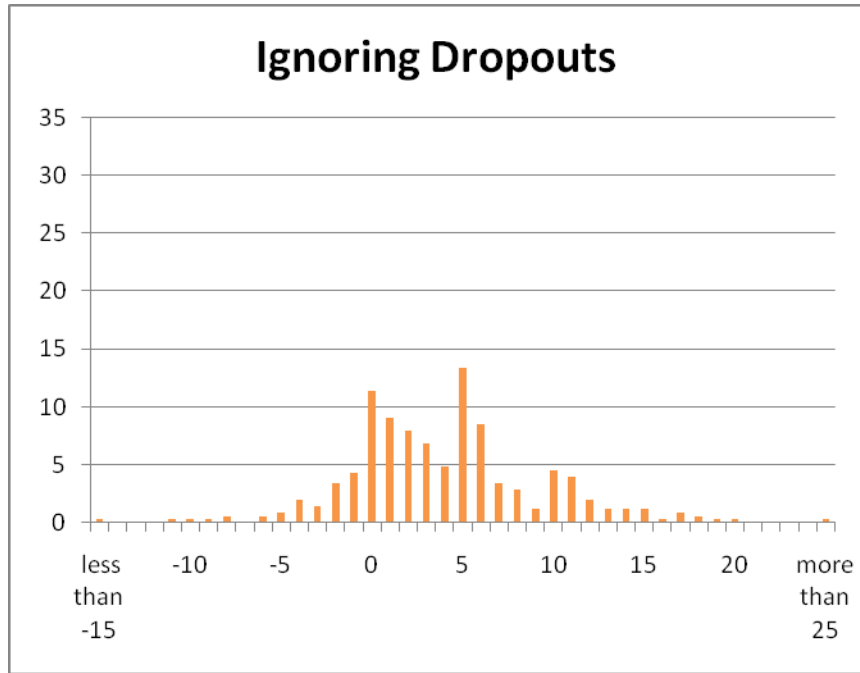


Figure 4-1: Distribution of Percent Weight Loss in the Control Group Quarter 1

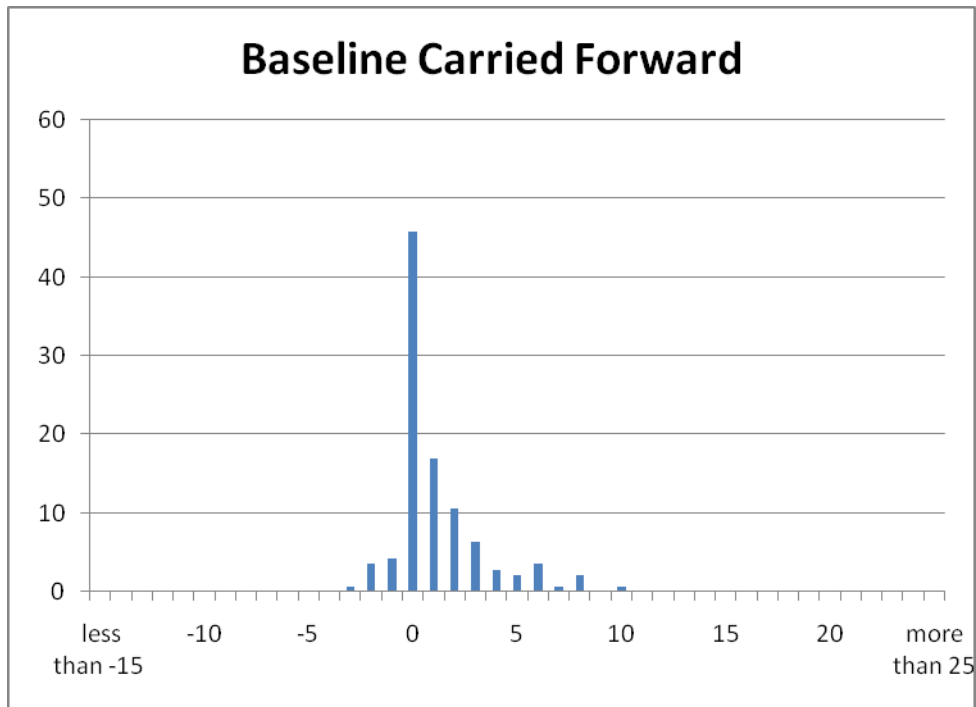
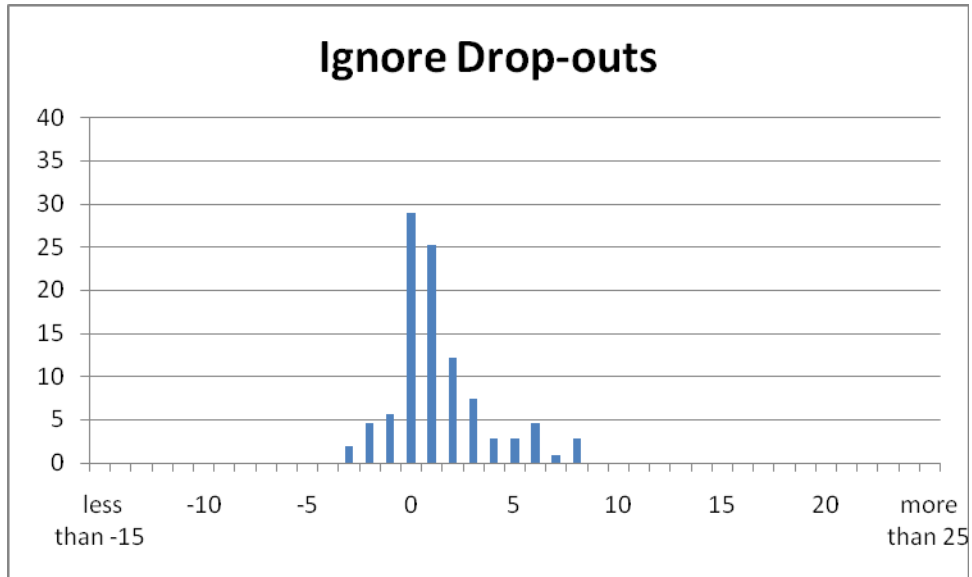


Figure 4-2: Distribution of Percent Weight Loss in the Control Group Quarter 2

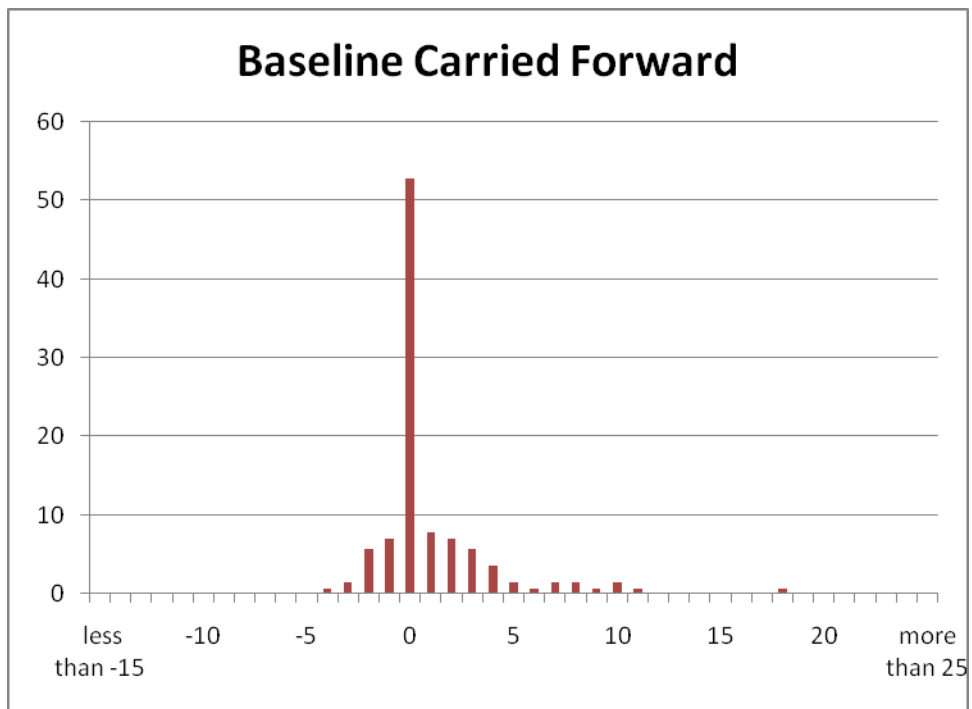
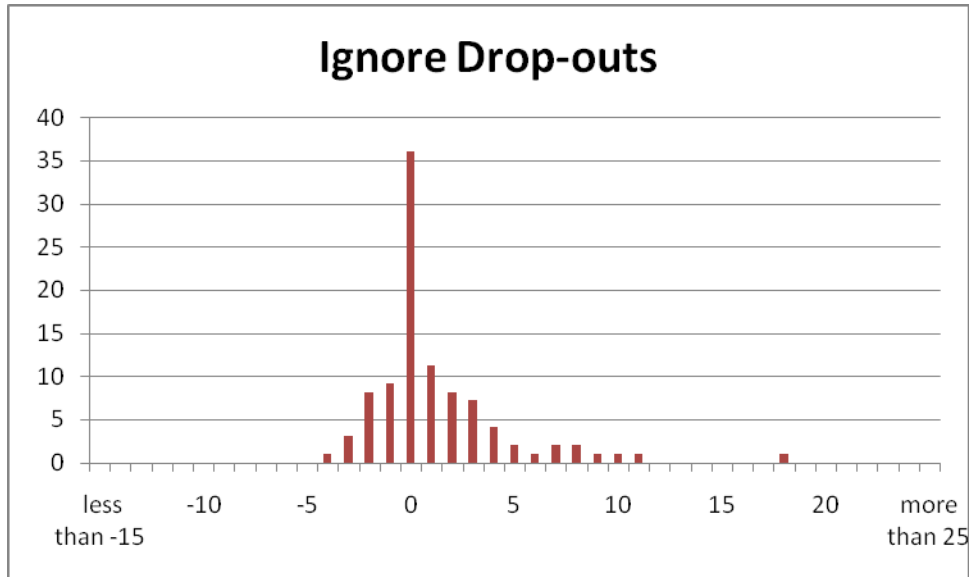


Figure 4-3: Distribution of Percent Weight Loss in the Control Group Quarter 3

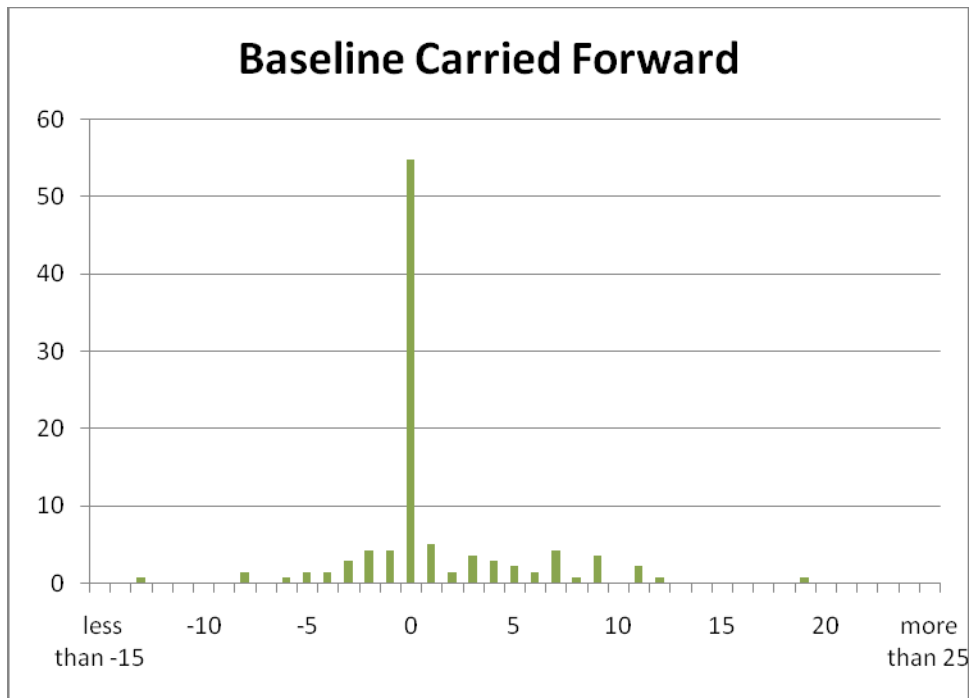
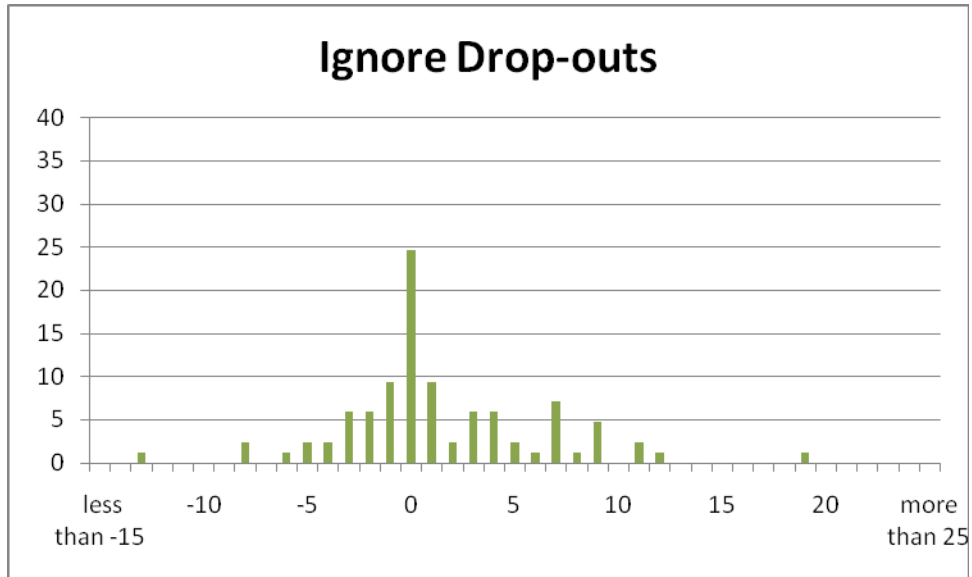
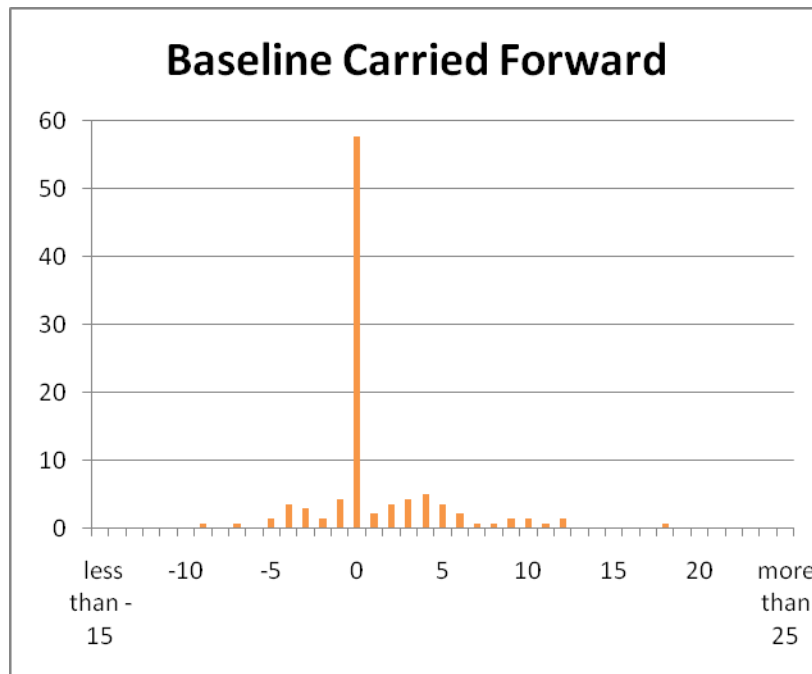
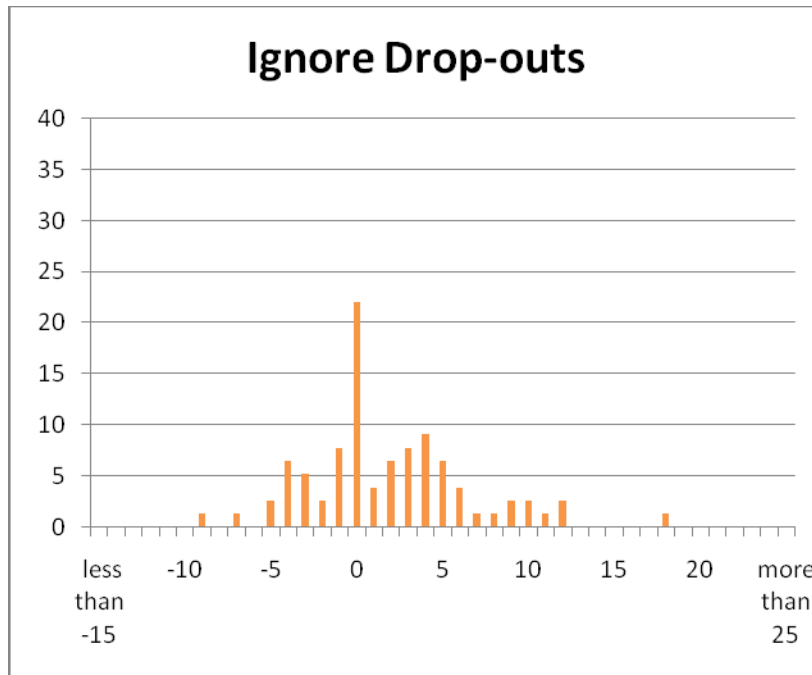


Figure 4-4: Distribution of Percent Weight Loss in the Control Group Quarter 4



**Table 1: Financial Rewards Based on Weight Loss
“Standard incentives”**

Weight Loss (as % of Baseline Weight)	Dollar Reward Per Month (Paid Quarterly)
1	1
2	2
3	3
4	4
5	5
10	10
15	15
20	25
25	35
30	50

**Table 2: Financial Rewards Based on Weight Loss
“Modified incentives”**

Weight Loss (as % of Baseline Weight)	Reward (Some Quarterly, Some Annual)
Greater than zero	Entered into quarterly drawing for gift certificates: ten \$50 gift cards each quarter and ten \$50 salon vouchers each quarter.
5	Complete reimbursement of monthly fees (12 * \$9.95), paid at end of year
10	Complete reimbursement of monthly fees (12 * \$9.95) plus \$100 bonus, paid at end of year
“Biggest loser” (as % of baseline) at worksite	\$250 gift certificate, awarded at end of year

Notes to Tables 1 and 2: Only participants with BMI over 25 (that is, those who are overweight or obese) are eligible to receive incentives. Moreover, people can only get incentives for weight loss down to a BMI of 25 – there is no financial incentive for anyone in the healthy weight (18.5 to 25) or underweight (<18.5) BMI categories to lose weight.

Table 3: Summary Statistics by Group

Variable	Standard incentives			Modified incentives			Control Group		
	Obs	Mean	Std. Dev	Obs	Mean	Std. Dev	Obs	Mean	Std.
Initial BMI	1374	32.9	6.34	835	32.9	6.08	142	31.3	5.58
Male	1374	0.159	0.37	835	0.214	0.41	142	0.345	0.477
Age	1374	45.1	10.42	835	42.7	8.89	142	43.9	10.44
Height	1374	65.5	3.42	835	66.2	3.41	142	66.7	4.138
Overweight (30>BMI>=25)	1374	0.410	0.49	835	0.380	0.49	142	0.514	0.502
Obese (40>BMI>=30)	1374	0.296	0.46	835	0.309	0.46	142	0.275	0.448
Morbidly Obese (BMI>=40)	1374	0.293	0.46	835	0.311	0.46	142	0.211	0.410
Foundation exercise regimen	1374	0.606	0.49	835	0.551	0.50	142	0.507	0.502
Intermediate exercise regimen	1374	0.334	0.47	835	0.375	0.48	142	0.408	0.493
Advanced exercise regimen	1374	0.060	0.24	835	0.074	0.26	142	0.085	0.279
Email open rate	555	45.4	35.44	834	50.3	34.85	142	28.0	32.04

Table 4: Enrollment Rates

	Control Group (1)	Standard Incentive Group (2)	Modified Incentive Group (3)	p-value (1) equals (2)	p-value (1) equals (3)	p-value (2) equals (3)
Mean	20.3%	28.4%	17.4%	0.381	0.726	0.223
(Std. Dev)	(10.0)	(11.3)	(17.4)			

Note: Enrollment rates are calculated by the fraction of those who enroll in the program by the total population of the work place.

Both the numerator and denominator include people of all BMI, even though financial rewards in the standard and modified incentives groups are paid only to those with BMI \geq 25; separate data are not available for those of BMI \geq 25.

Individuals with BMI $<$ 25 may enroll in the program, but receive no payouts.

Table 5: Cumulative Attrition, by Group and Quarter

Quarter	Standard Incentives	Modified Incentives	Control Group
1	46.4%	28.1%	23.2%
2	59.4%	33.7%	35.9%
3	70.7%	40.0%	41.5%
4	82.9%	57.8%	45.8%

Note: Data for control group only available for first three quarters; NA = Not Available.

Table 6: Results from Hazard Model of Attrition

Variable	Full Sample	Standard incentives	Modified incentives	Control Group
Age	0.964** (2.06)	0.961 (0.52)	0.965* (1.72)	0.937 (1.64)
Age Squared	1.000** (2.28)	1.000 (0.35)	1.000* (1.85)	1.000* (1.94)
Male	0.925 (1.23)	1.383 (1.33)	1.005 (0.06)	0.793* (1.86)
Obese	1.129** (2.13)	1.731** (1.99)	1.097 (1.35)	1.141 (1.17)
Morbidly Obese	1.140** (2.25)	1.515 (1.34)	1.061 (0.84)	1.304** (2.38)
Control Group	--			
Standard Incentive Group	2.299*** ^a (7.03)			
Modified incentives Group	1.112 (0.86)			
Intermediate Exercise Program	0.981 (0.37)	0.721 (1.32)	1.000 (0.01)	0.962 (0.39)
Advanced Exercise Program	0.839* (1.70)	0.742 (0.59)	0.800* (1.74)	0.937 (0.34)
Lagged Percent Weight Loss	0.9512*** (6.14)	1.001 (0.03)	0.987 (1.39)	0.869*** (9.14)
Sample Size	2351	1,374	835	142
Quarters Analyzed	4	4	4	4

Notes: Reference categories are overweight, standard incentives group, and Foundation exercise program. Table cells list hazard ratios and absolute value of z statistics in parentheses.

Table 7: Weight Loss in Pounds and Percent of Baseline Weight, by Group and Quarter

Quarter	Standard Incentives		Modified Incentives		Control Group	
	Ignoring Dropouts	Baseline Carry-Forward	Ignoring Dropouts	Baseline Carry-Forward	Ignoring Dropouts	Baseline Carry-Forward
1	3.5 (1.7%) Min = -12.6 Max = 56.4	1.9 (0.9%)	4.2 (2.0%) Min = -11.6 Max = 36	3.0 (1.5%)	3.6 (1.8%) Min = -6.0 Max = 25.2	2.8 (1.4%)
2	4.1 (2.0%) Min = -25.4 Max = 89.4	1.7 (0.8%)	4.8 (2.3%) Min = -18.4 Max = 52.8	3.2 (1.5%)	3.39 (1.8%) Min = -8.8 Max = 30.2	2.2 (1.1%)
3	5.6 (2.6%) Min = -56 Max = 109.2	1.7 (0.8%)	4.2 (2.0%) Min = -29.8 Max = 53.6	2.5 (1.2%)	3.62 (1.8%) Min = -22.2 Max = 32	2.1 (1.1%)
4	5.7 (2.7%) Min = -18.2 Max = 101.4	1.0 (0.5%)	8.6 (4.3%) Min = -30.6 Max = 61.2	3.6 (1.8%)	3.81 (2.00%) Min = -13.8 Max = 30.8	2.1 (1.1%)

Note: Data for control group only available for first three quarters; ; NA = Not Available.

A positive number indicates weight lost. A negative number (e.g. for the minimum weight loss) indicates weight gain.

**Table 8A: Results of Regression Model of Weight Loss in Pounds
Ignoring Those Who Drop Out (Completers Analysis)**

Variable	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Pooled
Age	0.169 [0.142]	0.161 [0.223]	0.069 [0.302]	0.414 [0.411]	0.186 [0.190]
Age ²	-0.001 [0.002]	-0.001 [0.003]	0 [0.003]	-0.004 [0.005]	-0.002 [0.002]
Male	2.498** [0.443]	2.752** [0.703]	3.254** [0.919]	2.858* [1.171]	2.851** [0.648]
Obese	0.544 [0.423]	0.936 [0.681]	0.467 [0.908]	1.696 [1.188]	0.828 [0.532]
Morbidly obese	1.567** [0.437]	1.786* [0.702]	2.511** [0.951]	2.580* [1.281]	1.989** [0.678]
Control Group	--	--	--	--	--
Standard incentives Group	0.218 [0.684]	0.926 [1.103]	2.229 ^a [1.425]	2.019 ^a [1.696]	1.326 [0.787]
Modified incentives Group	0.617 [0.700]	1.249 [1.115]	0.528 [1.419]	4.533** [1.645]	1.539 [0.796]
Intermediate exercise regime	0.586 [0.384]	0.418 [0.615]	-0.085 [0.831]	1.007 [1.099]	0.46 [0.520]
Advanced exercise regime	1.698* [0.676]	0.487 [1.110]	1.891 [1.449]	0.44 [1.899]	1.183 [0.877]
Quarter 2					0.576** [0.209]
Quarter 3					0.912** [0.334]
Quarter 4					3.182** [0.489]
Constant	-2.675 [3.109]	-2.731 [4.870]	-1.159 [6.515]	-7.981 [9.006]	-4.249 [4.089]
Observations	1445	1203	986	664	4298
R-squared	0.04	0.02	0.03	0.04	0.04

Standard errors in brackets

* significant at 10%, ** significant at 5%; *** significant at 1%

^a denotes significant difference between standard and modified incentive group at 5%

Note: Weight Loss is measured from baseline weight

Standard errors clustered at individual level in pooled regression.

**Table 8B: Results of Regression Model of Weight Loss in Pounds
Assuming Dropouts Return to Baseline Weight**

Variable	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Pooled
Age	0.177* [0.088]	0.133 [0.117]	0.153 [0.128]	0.193 [0.120]	0.164 [0.094]
Age ²	-0.002 [0.001]	-0.001 [0.001]	-0.002 [0.001]	-0.002 [0.001]	-0.002 [0.001]
Male	1.657** [0.297]	1.669** [0.393]	1.708** [0.432]	1.701** [0.403]	1.684** [0.370]
Obese	0.321 [0.275]	0.382 [0.363]	0.066 [0.400]	0.35 [0.373]	0.28 [0.266]
Morbidly obese	0.856** [0.282]	0.731* [0.372]	0.771 [0.409]	0.199 [0.382]	0.640* [0.316]
Control Group	--	--	--	--	--
Standard incentives Group	-0.608 ^a [0.488]	-0.255 ^a [0.645]	-0.208 [0.709]	-0.777 ^a [0.662]	-0.462 ^a [0.478]
Modified incentives Group	0.306 [0.506]	1.047 [0.668]	0.466 [0.735]	1.698* [0.686]	0.879 [0.524]
Intermediate exercise regime	0.432 [0.251]	0.295 [0.331]	-0.005 [0.364]	0.438 [0.340]	0.29 [0.258]
Advanced exercise regime	1.571** [0.481]	0.345 [0.635]	1.025 [0.698]	0.218 [0.652]	0.79 [0.521]
Quarter 2					-0.088 [0.108]
Quarter 3					-0.344* [0.142]
Quarter 4					-0.337* [0.158]
Constant	-2.631 [1.966]	-2.05 [2.596]	-2.3 [2.855]	-3.167 [2.666]	-2.345 [2.046]
Observations	2351	2351	2351	2351	9404
R-squared	0.03	0.02	0.01	0.04	0.02

Standard errors in brackets

* significant at 10%, ** significant at 5%; *** significant at 1%

^a denotes significant difference between standard and modified incentive group at 5%

Note: Weight Loss is measured from baseline weight

Standard errors clustered at individual level in pooled regression.

Table 9A: Results of Regression Model of Percent of Baseline Weight Lost Ignoring Those Who Drop Out (Completers Analysis)

Variable	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Pooled
Age	0.076 [0.069]	0.053 [0.107]	0.076 [0.141]	0.229 [0.196]	0.095 [0.091]
Age ²	-0.001 [0.001]	0 [0.001]	0 [0.002]	-0.002 [0.002]	-0.001 [0.001]
Male	0.798** [0.214]	0.793* [0.336]	1.067* [0.430]	0.713 [0.558]	0.862** [0.277]
Obese	-0.001 [0.204]	0.125 [0.326]	-0.093 [0.425]	0.353 [0.566]	0.071 [0.277]
Morbidly obese	0.016 [0.211]	-0.007 [0.336]	0.125 [0.446]	-0.251 [0.610]	-0.009 [0.296]
Control Group	--	--	--	--	--
Standard incentives Group	0.098 [0.330]	0.383 [0.528]	0.945 [0.667]	0.717 ^a [0.808]	0.553 [0.400]
Modified incentives Group	0.313 [0.338]	0.59 [0.534]	0.339 [0.665]	2.214** [0.784]	0.769 [0.408]
Intermediate exercise regime	0.364* [0.186]	0.306 [0.295]	-0.015 [0.389]	0.418 [0.524]	0.275 [0.255]
Advanced exercise regime	1.012** [0.326]	0.308 [0.531]	1.011 [0.679]	0.175 [0.904]	0.675 [0.424]
Quarter 2					0.274** [0.102]
Quarter 3					0.373* [0.156]
Quarter 4					1.524** [0.230]
Constant	-0.926 [1.502]	-0.366 [2.332]	-1.076 [3.051]	-3.673 [4.290]	-1.711 [1.962]
Observations	1445	1203	986	664	4298
R-squared	0.03	0.01	0.02	0.03	0.02

Standard errors in brackets

* significant at 10%, ** significant at 5%; *** significant at 1%

^a denotes significant difference between standard and modified incentive group at 5%

Note: Weight Loss is measured from baseline weight

Standard errors clustered at individual level in pooled regression.

**Table 9B: Results of Regression Model of Percent of Baseline Weight Lost
Assuming Dropouts Return to Baseline Weight**

Variable	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Pooled
Age	0.083 [0.043]	0.05 [0.056]	0.087 [0.060]	0.097 [0.057]	0.079 [0.044]
Age ²	-0.001 [0.000]	0 [0.001]	-0.001 [0.001]	-0.001 [0.001]	-0.001 [0.000]
Male	0.528** [0.143]	0.498** [0.188]	0.585** [0.202]	0.566** [0.192]	0.544** [0.158]
Obese	-0.019 [0.133]	-0.006 [0.174]	-0.123 [0.187]	0.007 [0.178]	-0.035 [0.139]
Morbidly obese	-0.035 [0.136]	-0.089 [0.178]	-0.064 [0.191]	-0.315 [0.182]	-0.126 [0.141]
Control Group	--	--	--	--	--
Standard incentives Group	-0.318 ^a [0.235]	-0.202 ^a [0.308]	-0.175 ^a [0.331]	-0.486 ^a [0.316]	-0.295 ^a [0.246]
Modified incentives Group	0.13 [0.244]	0.453 [0.320]	0.226 [0.343]	0.794* [0.327]	0.401 [0.268]
Intermediate exercise regime	0.276* [0.121]	0.216 [0.158]	0.033 [0.170]	0.199 [0.162]	0.181 [0.126]
Advanced exercise regime	0.931** [0.232]	0.256 [0.304]	0.574 [0.326]	0.144 [0.311]	0.476 [0.251]
Quarter 2					-0.045 [0.053]
Quarter 3					-0.195** [0.068]
Quarter 4					-0.168* [0.074]
Constant	-1.029 [0.948]	-0.405 [1.242]	-1.17 [1.333]	-1.302 [1.271]	-0.874 [0.969]
Observations	2351	2351	2351	2351	9404
R-squared	0.03	0.02	0.01	0.04	0.02

Standard errors in brackets

* significant at 10%, ** significant at 5%; *** significant at 1%

^a denotes significant difference between standard and modified incentive group at 5%

Note: Weight Loss is measured from baseline weight

Standard errors clustered at individual level in pooled regression.

Table 10: Percent of Respondents Losing 5% of Baseline Weight, by Group and Quarter

Quarter	Standard Incentives		Modified Incentives		Control Group	
	Ignoring Dropouts	Baseline Carry-Forward	Ignoring Dropouts	Baseline Carry-Forward	Ignoring Dropouts	Baseline Carry-Forward
1	14.1%	7.6%	16.5%	11.9%	11.9%	9.2%
2	19.4%	7.9%	23.8%	15.8%	13.2%	8.4%
3	24.9%	7.3%	23.0%	13.8%	26.5%	15.5%
4	24.3%	4.1%	45.7%	19.3%	23.4%	12.7%

Note: Data for control group only available for first three quarters; NA = Not Available.

**Table 11A: Results of Probit Model of Probability of 5% Weight Loss
Ignoring Those Who Drop Out (Completers Analysis)
(Marginal Effects are Presented)**

Variable	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Pooled
Age	0.002 (0.008)	-0.008 (0.009)	-0.005 (0.011)	-0.005 (0.015)	-0.003 (0.008)
Age ²	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Male	0.044* (0.026)	0.029 (0.031)	0.072** (0.036)	0.046 (0.044)	0.051** (0.025)
Obese	0.004 (0.023)	0.012 (0.029)	-0.012 (0.032)	0.006 (0.043)	0.000 (-0.023)
Morbidly Obese	0.011 (0.024)	-0.014 (0.029)	-0.008 (0.034)	-0.082* (0.044)	-0.020 (0.023)
Control	--	--	--	--	--
Standard Incentive	0.033 (0.041)	0.082 (0.056)	-0.002 (0.050)	0.022 ^a (0.065)	0.024 ^a (0.039)
Modified Incentive	0.048 (0.044)	0.119** (0.059)	-0.034** (0.048)	0.227*** (0.068)	0.075* (0.042)
Intermediate	0.036* (0.023)	0.040 (0.027)	0.014 (0.031)	0.037 (0.041)	0.033 (0.022)
Advanced	0.105** (0.046)	0.045 (0.051)	0.090* (0.058)	0.071 (0.071)	0.080** (0.038)

* significant at 10%, ** significant at 5%, *** significant at 1%,

^a denotes significant difference between standard and modified incentive group at 5%
Significance is measured using probit coefficients and standard errors

Standard errors of marginal effects in brackets, clustered at individual level in pooled regressions

**Table 11B: Results of Probit Model of Probability of 5% Weight Loss
Baseline Carry-Forward
(Marginal Effects are Presented)**

Variable	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Pooled
Age	0.004 (0.005)	-0.002 (.005)	0.002 (0.005)	0.002 (0.005)	0.001 (0.004)
Age ²	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Male	0.027* (0.017)	0.018 (0.017)	0.037** (0.018)	0.040*** (0.017)	0.031** (0.013)
Obese	0.000 (0.014)	-0.003 (0.015)	-0.015 (0.014)	-0.009 (0.013)	-0.007 (0.011)
Morbidly Obese	0.003 (0.015)	-0.015 (0.015)	-0.015 (0.014)	-0.046*** (0.012)	-0.018 (0.011)
Control	--	--	--	--	--
Standard Incentive	-0.009 ^a (0.025)	-0.001 ^a (0.029)	-0.067** ^a (0.019)	-0.085*** ^a (0.016)	-0.041** ^a (0.017)
Modified Incentive	0.025 (0.030)	0.073** (.038)	-0.011 (0.023)	0.053** (0.029)	0.033 (0.023)
Intermediate	0.027 (0.015)	0.028** (0.015)	0.011 (0.014)	0.020 (0.014)	0.022** (0.011)
Advanced	0.096*** (0.035)	0.040 (0.031)	0.055** (0.031)	0.038 (0.028)	0.058*** (0.023)

* significant at 10%, ** significant at 5%, *** significant at 1%,

^a denotes significant difference between standard and modified incentive group at 5%

Significance is measured using probit coefficients and standard errors

Standard errors of marginal effects in brackets, clustered at individual level in pooled regressions

Table 12: Percent of Respondents Losing 10% of Baseline Weight, by Group and Quarter

Quarter	Standard Incentives		Modified Incentives		Control Group	
	Ignoring Dropouts	Baseline Carry-Forward	Ignoring Dropouts	Baseline Carry-Forward	Ignoring Dropouts	Baseline Carry-Forward
1	1.9%	1.0%	2.7%	1.9%	0.9%	0.7%
2	4.8%	2.0%	6.5%	4.3%	4.4%	2.8%
3	9.0%	2.6%	6.8%	4.1%	6.0%	3.5%
4	9.8%	1.7%	16.5%	6.9%	7.8%	4.2%

Note: Data for control group only available for first three quarters; NA = Not Available

**Table 13A: Results of Probit Model of Probability of 10% Weight Loss
Ignoring Those Who Drop Out (Completers Analysis)
(Marginal Effects are Presented)**

Variable	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Pooled
Age	0.007* (0.004)	-0.001 (0.005)	-0.008 (0.006)	-0.006 (0.010)	-0.002 (0.004)
Age ²	-0.000* (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Male	0.026** (0.014)	0.033* (0.021)	0.003 (0.021)	0.054* (0.035)	0.030** (0.016)
Obese	0.018* (0.013)	0.014 (0.018)	0.022 (0.023)	-0.003 (0.030)	0.012 (0.014)
Morbidly Obese	0.013 (0.014)	0.000 (0.017)	0.009 (0.023)	-0.047 (0.028)	-0.004 (0.014)
Control	--	--	--	--	--
Standard Incentive	0.022 (0.031)	0.012 (0.031)	0.028 (0.039)	0.041 (0.057)	0.014 (0.024)
Modified Incentive	0.019 (0.030)	0.022 (0.033)	0.008 (0.035)	0.090* (0.058)	0.027 (0.026)
Intermediate	0.008 (0.010)	0.006 (0.015)	-0.013 (0.017)	0.005 (0.029)	0.002 (0.012)
Advanced	0.016 (0.020)	-0.017 (0.022)	0.001 (0.032)	-0.027 (0.044)	-0.002 (0.017)

* significant at 10%, ** significant at 5%, *** significant at 1%,

^a denotes significant difference between standard and modified incentive group at 5%
Significance is measured using probit coefficients and standard errors

Standard errors of marginal effects in brackets, clustered at individual level in pooled regressions

**Table 13B: Results of Probit Model of Probability of 10% Weight Loss
Baseline Carry-Forward
(Marginal Effects are Presented)**

Variable	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Pooled
Age	0.005* (0.003)	0.000 (0.003)	-0.002 (0.003)	0.000 (0.003)	0.000 (0.002)
Age ²	-0.000* (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Male	0.016** (0.009)	0.019** (0.011)	0.004 (0.010)	0.027** (0.013)	0.017*** (0.008)
Obese	0.011* (0.008)	0.005 (0.009)	0.005 (0.009)	-0.005 (0.008)	0.003 (0.006)
Morbidly Obese	0.007 (0.008)	-0.002 (0.008)	0.000 (0.009)	-0.021** (0.007)	-0.005 (0.006)
Control	--	--	--	--	--
Standard Incentive	0.009 (0.017)	-0.005 ^a (0.014)	-0.009 (0.014)	-0.023 ^a (0.012)	-0.007 ^a (0.009)
Modified Incentive	0.012 (0.020)	0.015 (0.019)	0.005 (0.017)	0.024 (0.020)	0.014 (0.013)
Intermediate	0.005 (0.006)	0.004 (0.008)	-0.004 (0.008)	0.005 (0.009)	0.002 (0.006)
Advanced	0.013 (0.014)	-0.007 (0.012)	0.007 (0.016)	-0.005 (0.014)	0.002 (0.009)

* significant at 10%, ** significant at 5%, *** significant at 1%,

^a denotes significant difference between standard and modified incentive group at 5%

Significance is measured using probit coefficients and standard errors

Standard errors of marginal effects in brackets, clustered at individual level in pooled regressions

Appendix Table 1:
Previous Literature on Financial Incentives for Weight Loss

Study	Study Design	Intervention and Incentives	Sample Size and Population	Duration	Weight Loss	Attrition
Volpp et al. (2008)	Randomized controlled trial	<p>3 groups:</p> <p>Deposits contract of \$0-\$3 / day matched 1:1.</p> <p>Lottery for daily prize with $E[V]=\\$3$.</p> <p>Self-reported daily weight.</p> <p>\$20 for monthly weigh- in, unconditionally</p>	<p>N=57 (19 in each of 3 groups)</p> <p>Patients at Philadelphia Veterans Affairs Medical Center with BMI 30-40</p>	16 weeks	<p>Mean weight loss:</p> <p>Lottery: 13.1 lbs</p> <p>Deposit contracts: 14.0</p> <p>Control: 4.0 lbs.</p>	8.4%
Burger and Lynham (2008) working paper	Opportunistic data from William Hill betting agency for 1993-2006	<p>Maximum bet of \$65. William Hill offered odds ranging from 5:1 to 50:1; potential payoff averaged \$1,926.</p> <p>Average duration of bet is 8 months, weight to be lost ranges from 28-168 lbs.</p> <p>Each bettor weighed at start and end of bet by physician.</p> <p>No control group.</p>	<p>N=51</p> <p>Self-selected members of British population.</p>	Average of 8 months	Approximately 80% of people betting on their weight loss lose the bet	Approximately 80% of people betting on their weight loss lose the bet
Finkelstein et al. (2007)	Randomized trial, no control group	<p>Three groups:</p> <p>Back loaded: \$0 at 3 months, \$14 per % point lost at 6 months</p> <p>Front loaded: \$14 per % point lost at 3 months, \$0 at 6 months</p> <p>Steady payment: \$7 per % point lost at</p>	<p>N=207 (72 in Back Loaded, 64 in Front Loaded, 71 in Steady Payment)</p> <p>Overweight and obese employees at one university and 3</p>	6 months	<p>Mean weight loss 3 months:</p> <p>2 lbs for Back Loaded, 4.7 lbs for Front Loaded, 3 lbs for Steady Payment</p> <p>Mean weight loss at 6</p>	<p>54% in Back Loaded, 45% in Front Loaded, 31% in Steady Payment</p>

		both 3 months and 6 months Weigh-ins at 3 months and 6 months. Incentives only up to 10% weight loss (\$140)	community colleges in NC		months not significantly different from zero	
Butsch et al. (2007)	Sequential control-intervention, not randomized	Treatment group eligible for 50% reimbursement of enrollment fee (\$150 of \$300) if lose 6% of initial weight and attend 10 of 12 group sessions Control group was not eligible for reimbursement	N=401 (241 intervention of which 59 enrolled, 160 control of which 40 enrolled) Participants in Univ. Alabama at Birmingham EatRight Lifestyle Program BMI 30 and over.	12 weeks	Mean weight loss: 2.25% in control group, 3.27% in intervention group; difference not statistically significant	Not stated
Hubbert et al. (2003)	Propensity score matching of 4 controls to each member of intervention group	Treatment group eligible for 50% of cost of program fees (\$150 of \$300) if lose 6% of initial weight and attend 10 of 12 group sessions. Control group was not eligible for reimbursement	N=125: 25 in intervention group, 100 in control group Participants in Univ. Alabama at Birmingham EatRight Lifestyle Program and members of UAB-owned HMO BMI 30 and over.	12 weeks	Mean weight loss: 7.3 kg (6.1%) in intervention group, 4.0 kg (3.9%) in control group; both differences are statistically significant	Not stated
Jeffery, Forster, et al. (1993)	Block-randomized controlled experiment (worksites randomized)	Worksites divided evenly between treatment and control groups. Treatment (Healthy Worker Project) consisted of health education classes and payroll deductions that served as	32 worksites in Minneapolis / St. Paul metropolitan area. Of 10,000 employees in treatment worksites,	2 years	No treatment effect was found for weight. In cohort survey, average change in BMI was 0.08 units for	No attrition of worksites.

		<p>bonds – refunded if achieve weight loss goals or donated to charity otherwise.</p> <p>Goals chosen by employee and ranged from minimum of 0 lb and maximum of 1% body weight loss each week.</p> <p>Participants chose amount of payroll deduction (minimum of \$5 biweekly).</p> <p>200 employees surveyed at baseline and again after 2 years (cohort). Another 200 employees surveyed after 2 years (cross-section).</p> <p>Weight self-reported but corrected for reporting error.</p>	2,041 employees participated in weight control program.		<p>control group, -0.02 units for treatment group; not statistically significant.</p> <p>In cross-sectional survey, average change in BMI was -0.05 in both the treatment and control groups.</p>	
Jeffery, Wing, et al. (1993)	Randomized controlled experiment	<p>Five groups: 1) control; 2) standard behavioral therapy (SBT); 3) SBT plus food provision; 4) SBT plus incentives; 5) SBT plus food provision plus incentives.</p> <p>Weekly incentives: \$0 if gained weight, \$2.50 if did not gain weight; \$12.50 if weight loss was 50% of goal, \$25 if weight loss reached goal.</p> <p>Weight-loss goals could be either 14, 18, or 23 kg during course of program.</p> <p>Weight measured at baseline, 6, 12, and 18 months. There were also optional weekly weigh-ins.</p>	<p>N=202 men and women from Pittsburgh and Minneapolis-St. Paul, of which 40 to 41 were in each of the 5 groups.</p> <p>Had to be 14-32 kg overweight</p>	18 months	No effect of financial incentives or the interaction of financial incentives with food provision	11% attrition at 6 months, 13% at 12 months, 15% at 18 months
Jeffery et al. (1990)	Randomized experiment	Two groups: 1) offered a weight control newsletter program for price of \$5; 2) offered the same program for	N=1,304 residents of Bloomington Minnesota:	6 months	Weight loss averaged about 4 lbs for \$5 program and 8 lbs for	3.8% did not return survey

		<p>free but requiring a \$60 deposit that would be refunded based on (proportional to) success in weight loss.</p> <p>Individuals chose weight loss goals of not more than 4 lb a month.</p> <p>Weight self-reported (questionnaire, telephone survey). For subset of respondents, validation of self-report through measurement of weight.</p>	<p>1,190 in the \$5 newsletter program group and 114 in newsletter plus incentive program group</p>		<p>incentive program.</p>	
Kramer et al. (1986)	Randomized controlled experiment	<p>Three groups: 1) monthly financial contingencies for weight maintenance; 2) monthly financial contingencies for participation in training sessions to solidify behavioral changes; 3) no treatment.</p> <p>\$120 deposit. For each of 12 sessions not attended, participant forfeited \$10. Refund also withheld if weighed more than “baseline” (post-first-treatment) weight. Withheld refunds (forfeited moneys) were distributed among those who were at or below “baseline” weight at final session.</p> <p>Weight measured at “baseline” and at one year.</p>	<p>N=85 individuals who had already lost 10% or more of their body weight through a 15-week weight-loss program.</p>	1 year	<p>Incentives had no impact on weight maintenance / amount of weight regained. Average weight regain: 10.3 lbs in control group, 11.9 lbs. in group with incentives.</p>	<p>6 of 28 (21%) of the incentives group refused to attend final weigh-in. They self-reported weight, and 5 lbs was added to account for under-reporting.</p>
Jeffery et al. (1984)	Randomized controlled experiment	<p>Three groups: 1) regular contract, 2) difficulty-grade contract; 3) no contract (control)</p> <p>All deposited \$150. Immediately refunded to control group. Regular contract group received \$30 for each 5-lb. increment of weight loss</p>	<p>N=113</p> <p>Roughly half recruited from population sample and the other half from newspaper advertisements</p>	15 weeks	<p>Average weight loss: 26.2 lbs (12.8%) in difficulty-grade contract 21.7 lbs (10.8%) in regular contract 17.7 lbs (8.5%) in control group</p>	<p>11 subjects (10%) refused to attend final weigh-in. They self-reported weight, and 5</p>

		Difficult-grade contract group received \$5 for first 5 lbs lost, \$10 for second, \$20 for third, \$40 for fourth, and \$75 for fifth.				lbs was added to account for under-reporting.
Jeffery et al. (1983)	Randomized experiment	<p>Six treatment groups: 3 levels of deposit (\$30, \$150, \$300) times two types of payoff criteria: individual weight loss or mean group weight loss.</p> <p>All received 15-week behaviorally oriented program. Goal was 30 pounds lost.</p> <p>Cash refunds per week at rate of \$1, \$5, or \$10 per pound up to 2 pounds per week.</p> <p>Monies not refunded for weight loss by end of program were distributed equally among those who achieved the 30-pound weight loss goal.</p> <p>Participants were weighed weekly.</p>	<p>N=89</p> <p>Men in the Minneapolis area with self-reported weight at least 30 pounds above the ideal.</p>	15 weeks	<p>Individuals rewarded for group performance lost on average 5 lbs. more weight. This difference was maintained over 1 year follow-up.</p> <p>No significant effects of contract size.</p>	None
Coates et al. (1982)	Randomized experiment	<p>Four treatment groups: 2 incentivized behaviors (weight loss or decrease in calorie consumption) by 2 frequencies of therapeutic contact (5 times or 1 time per week)</p> <p>Deposits were equal to 15 weeks' allowance or 50% of earnings from part-time work; amounts varied from \$15-\$240 (mean=\$67.75). Source of payment: parents (51.5%), subjects (39.4%), shared (9.1%).</p>	<p>N=36</p> <p>Adolescents at least 10% above average weight-for-height.</p>	15 weeks	<p>The treatment group receiving rewards for weight loss and coming to the clinic 5 times per week was the only group to significantly reduce the percent overweight. Treatment effects maintained over a 6-month follow-up period.</p>	None

		<p>Weight loss goal was 1 lb. per week, or caloric reduction necessary to lose 1 lb. per week. Monetary reward was delivered either once per week or once per week at treatment center.</p> <p>Weighed at each clinic visit. Food records checked</p>			<p>Significant correlation between initial monetary deposit and percent overweight lost.</p> <p>No significant difference based on whether parents or subject paid the deposit</p>	
Jeffery et al. (1978)	Randomized controlled experiment.	<p>Three treatment groups: deposits were returned contingent on either attendance, calorie restriction, or weight loss. Also a control group.</p> <p>Each of the three treatment groups deposited \$200. One group paid \$20 for losing 2 lbs. per week. Another paid \$20 for calorie restrictions calculated to cause loss of 2 lbs. per week. Third group paid \$20 for weekly attendance.</p>	<p>N=31</p> <p>Respondents to newspaper advertisement for people who need to lose 50 lbs. or more.</p>	10 weeks	<p>Groups rewarded for weight loss or calorie reductions lost an average of 20 lbs, significantly more weight loss than either the group rewarded for attendance (8.6 lbs) or the control group (12.4 lbs).</p>	4 of 7 in control group quit.
Mann (1972)	Single-subject reversal design	<p>Subjects deposited a large number of valuables (e.g. money, jewelry, medals) with the researcher and signed a Contingency Contract allowing the researcher to switch them from treatment to control conditions, with the treatment being valuables being either returned or forfeited based on weight loss.</p> <p>One valuable was returned for each 2 lb weight loss over a 2-week period.</p>	<p>N=8</p> <p>Respondents to newspaper advertisement. All agreed to lose 25 pounds or more and had physician approval.</p>	Durations of treatments varied; total study ran at least 400 days	Average weekly weight loss of 1.6 to 1.7 pounds during treatment, regain of 1.4 pounds per week when incentives removed.	None

		Subjects weighed every Monday, Wednesday, and Friday.				
--	--	--	--	--	--	--