



# Examining Ways to Improve Weight Control Programs' Population Reach and Representativeness: A Discrete Choice Experiment of Financial Incentives

Wen You<sup>1</sup> · Yuan Yuan<sup>2</sup> · Kevin J. Boyle<sup>3</sup> · Tzeyu L. Michaud<sup>4</sup> · Chris Parmeter<sup>5</sup> · Richard W. Seidel<sup>6</sup> · Paul A. Estabrooks<sup>4</sup>

Accepted: 12 October 2021 / Published online: 10 November 2021  
© The Author(s) 2021

## Abstract

**Background** Both theoretical and empirical evidence supports the potential of modest financial incentives to increase the reach of evidence-based weight control programs. However, few studies exist that examine the best incentive design for achieving the highest reach and representativeness at the lowest cost and whether or not incentive designs may be valued differentially by subgroups that experience obesity-related health disparities.

**Methods** A discrete choice experiment was conducted ( $n = 1232$  participants; over 90% of them were overweight/obese) to collect stated preference towards different financial incentive attributes, including reward amount, program location, reward contingency, and payment form and frequency. Mixed logit and conditional logit models were used to determine overall and subgroup preference ranking of attributes. Using the National Health and Nutrition Examination Survey data sample weights and the estimated models, we predicted US nationally representative participation rates by subgroups and examined the effect of offering more than one incentive design. External validity was checked by using a completed cluster randomized control trial.

**Results** There were significant subgroup differences in preference toward incentive attributes. There was also a sizable negative response to larger incentive amounts among African Americans, suggesting that higher amounts would reduce participation from this population. We also find that offering participants a menu of incentive designs to choose from would increase reach more than offering higher reward amounts.

**Conclusions** We confirmed the existence of preference heterogeneity and the importance of subgroup-targeted incentive designs in any evidence-based weight control program to maximize population reach and reduce health disparities.

## Key Points for Decision Makers

Larger incentive rewards can reduce program participation from certain subgroups such as African Americans.

Financial incentive design is a product with multiple interdependent components; therefore, the best fit design should not focus only on reward amount.

Offering a menu of financial incentive conditions for participants to choose from will increase program participation more than offering larger reward amounts.

## 1 Background

The obesity epidemic and its related negative impact on public health, both on quality of life and the national economy, have given rise to an increase in workplace, healthcare, and community efforts to move evidence-based interventions (EBIs) into typical practice [1, 2]. However, there is limited evidence that available EBIs are effectively engaging subgroups of the population that experience disparities in obesity, and obesity-related outcomes, for the duration necessary to achieve a clinically meaningful weight loss (i.e.,  $\geq 5\%$  of initial weight) [3]. Some subgroups, such as African Americans, Latino, and lower socioeconomic status groups [4], have been shown to be more likely to experience obesity but less likely to engage in weight control EBIs [3, 5]. Understanding how best to reach populations that experience obesity-related health disparities in real-world settings is a

✉ Wen You  
wenyou@virginia.edu

Extended author information available on the last page of the article

key area of need in the field of dissemination and implementation science. Furthermore, it is a vital step towards achieving the full public health potential of weight-loss EBIs and eliminating obesity-related health disparities [3].

There is emerging evidence, and a strong theoretical rationale, supporting the use of financial incentives as an implementation strategy to improve the reach of EBIs focused on weight loss [6–15]. The use of incentives is also endorsed by the WHO [16] as an important component of behavioral weight control programs and was included in the US Patient Protection and Affordable Health Care Act (ACA). Starting in 2014, Section 2705 of the ACA allowed employers to use up to 30% of the total amount of health insurance premiums to provide outcome-based wellness incentives [17]. Nevertheless, it appears as though the benefit of using financial incentives to engage participants in wellness and weight loss programs may not achieve the desired outcomes relative to program reach [18–20]. For example, participation rates in weight control programs do not align with rates that one would expect based on economic theory—that is, applying a financial incentive should engage a larger proportion of the intended audience if well calibrated [19]. The RAND Health report showed that the uptake of worksite wellness programs is limited with eligible employee participation rates ranging from 21% (for fitness programs) to 7% (for smoking cessation programs). The weight loss program participation rate, at 10%, was the second lowest rate identified in the report [21].

One of the largest challenges to understand the use of financial incentives to improve the reach of weight loss programs is the limited variety of incentive designs used within and across studies. It is likely that heterogeneity exists in preferences towards different financial incentive attributes such as magnitude, form, frequency, and contingency [20]. Consider an adult man and woman who are obese and from different racial groups: when presented with an opportunity to participate in a weight loss program, each considers the competing demands, interests, and values relative to budget constraints. These constraints are likely different for the two individuals due to different experiences and motivations that are related to their respective backgrounds and characteristics (e.g., health literacy levels and information exposure differences). Changes to a healthier diet could be perceived as more costly to one person, but not the other. Changes could also be perceived differently for these individuals. One may see participation as negatively impacting satisfaction with a new status quo due to a shift from more preferred, less healthful foods to less preferred, more healthful foods—which is a disincentive to participation. The addition of a financial incentive to the equation may offset the time and psychological costs faced (e.g., limiting food she enjoys, initiating new physical activities) by

each. However, too high of an incentive may lead to health behaviors that are only sustained as long as the incentive is provided [22, 23]. Modest incentives are intended to give temporary extrinsic motivation that is intended to give way to intrinsic motivation as a participant begins to realize the benefits of weight loss. The amount, type, timing, and location of incentives needed will vary across these individuals. Just considering the amount of incentives, one of the hypothetical people in our example may be more satisfied with the status quo (i.e., not initially motivated to lose weight) and would need a higher amount. The other may be dissatisfied with the status quo and therefore more motivated to change. This person may not need any incentives, but the application of an incentive may have the potential to disrupt initial intrinsic motivation.

While there is strong justification to examine the potential heterogeneity of preferences for incentive models to improve the reach of evidence-based weight loss programs, most of the research in this area has examined different incentive amounts on program effectiveness and maintenance with only a limited number of studies examining incentives as an implementation strategy to improve reach (i.e., program enrollment/participation rate). Furthermore, the current literature found mixed evidence on how financial incentives may effectively improve program reach [24–27]. The mixed evidence may be due to the fact that studies did not thoroughly consider incentive constructs and led to the objective of this study—to inform better incentive designs and implementation through understanding the preferences of individuals towards different financial incentive attributes.

## 2 Experimental Method and Implementation

The development of tailored incentive designs for subgroups of the population relies on the understanding of target population preferences towards different incentive attributes. We used a discrete choice experiment (DCE) to understand public preferences towards different financial incentive attributes. The DCE has been used widely in health economics and the studies that examine preferences towards weight loss program attributes are especially relevant to this study [28–30]. The application of choice experiment allows us to examine a wider range of incentive attributes (i.e., more and broader ranges of levels within each attribute), overcoming the limitations of actual field experiments (i.e., limited scope, limited representativeness, and limited incentive designs). It further allows larger sample sizes for procuring sufficient statistical power and oversampling of those at-risk groups that were often difficult to recruit in weight loss programs and randomized control trials.

## 2.1 Survey Instrument Development

To serve the purposes of customization and targeted design, care was taken to (1) identify and describe the incentive attributes that are plausible and relevant to those at-risk subgroups; and (2) adjust survey questionnaires' readability to accommodate low literacy levels in order to ensure survey accessibility and data quality (i.e., minimize data noises caused by comprehension of the survey itself). At-risk subgroups were identified a priori and were based on the literature documenting disparities in obesity-related health outcomes and program participation for these groups (e.g., gender, race, and socio-economic status) [4]. The research team involved stakeholders in the survey development through a 2-h listening session with 22 program assistants and nutrition outreach instructors from the Expanded Food and Nutrition Education Program (EFNEP). Those educators worked with low-income populations who experience a high rate of overweight and obesity. Based on the information gained from the listening session, the initial questionnaire was drafted and subsequently a total of four focus groups were held, approximately 1 month apart, to test, refine, and pilot the survey instruments. Focus group participants were recruited from EFNEP and Carilion Clinic (a health-care organization in southwest Virginia partnering with the research team), were overweight or obese, and the majority of them were low income (i.e., Medicaid eligible).<sup>1</sup> The list of financial incentive attributes and their levels were formed and finalized during the process. This use of qualitative evidence in survey design follows recommended practice [31]. The initial formation of the list was also aided by an extensive literature review and a pilot study of the project. The final list of incentive attributes and their levels used in our DCE are presented in Table 1.<sup>2</sup> The Appendix Figs. 5 and 6 show the survey description of the weight control program and incentive schemes.

A total of 96 choice sets (each containing two incentive designs comprising unique combinations of incentive attributes and an opt-out option) were chosen based

<sup>1</sup> A total of four focus groups were held between April 2011 and July 2011. The numbers of participants in the four focus groups are 9, 9, 12, and 10. The improvements made after those focus groups were (a) readability improvement to the descriptions of the weight control program and incentives: pictures and examples were used and layout and flow of questions were updated; (b) finalized incentive attributes and levels to be more relevant to target respondents (e.g., we dropped grocery card as a form of payment after learning that it was perceived to be the same as cash).

<sup>2</sup> We included a \$0 amount in the Reward amount attribute level list to reflect the fact that the incentive has more dimensions than the magnitude itself (i.e., dollar amount). When financial reward amount is \$0, the program location, weekly weight check-in, and turning-in of diet and physical records are still relevant and are non-financial incentives.

**Table 1** Attributes and attribute levels used in constructing the financial incentive

Attributes	Attribute levels
Reward amount	\$0 \$48 \$96 \$216 \$384 \$576
Payment form	Cash Pre-paid gym pass (gym) *Health debit card for doctor's visits, prescriptions, and other medical expenses (medical) Debit card (debit)
Payment frequency	*Once at end of program (once) Quarterly Monthly Weekly
Program location	*Clinic Workplace Community center Church
Reward condition	Losing 2 lbs (weight) *Attending weekly weight checks (attendance) Turning in records of diet and exercise (compliance) Attending weekly weight checks and turning in records of diet and exercise (combined)

Abbreviated terms used in the text are in parentheses. Baseline levels for categorical attributes are denoted by \*. Reward amount is defined as the total possible payment over the duration of the 6-month intervention

on the final list of attributes and associated levels using D-efficiency criterion and the design was done using software package *Ngene*. The D-efficiency design has been shown to provide more statistically efficient parameter estimates, on average, than other designs that are less D-efficient [32]. Each survey includes four choice sets. The choice sets generally can range from 1 to 18 and the decision should be informed by contextual knowledge and the trade-off between respondents' burden and correlation within and between subjects [33]. We chose to ask each participant to respond to four choice sets based on what we learned in the focus group interviews related to participant burden considering our tasks' numeric nature. The Appendix Fig. 7 presents an example of the choice question (see ESM).

## 2.2 Survey Implementation

Survey participants were identified through the electronic patient database of our research team partner, Carilion

Clinic.<sup>3</sup> A total of 7554 adults (aged 18 years or older) with the most recent clinic record of non-missing BMI (used to identify and oversample those with BMI  $\geq 25$ ) and non-missing mailing information were identified. Those patients were mailed introductory letters, signed by the Carilion Clinic primary care branch director, describing the study goals, mail survey expected length of completion time (15–25 min), and informing about the upcoming phone recruitment and the standard confidentiality statement. The phone recruitment and mail survey implementation were all conducted by Virginia Tech Survey Research Center following standard recruitment rules in 2012. Among those 7554 identified individuals, about 6657 (88%) of them had valid phone numbers on record or were reachable by the recruiters over the phone. The phone recruitment resulted in a total of 2737 individuals (41% response rate) who consented to participation in the mail survey. Those individuals later received the survey instruments via mail along with US\$2 as a thank you for participation. Two to three weeks after the first round of survey mail-out, those who had not responded were mailed the same survey a second time. A total of 1232 participants completed the survey. During the recruitment process, specific emphasis was placed on oversampling low-income (identified by the Medicaid eligibility information in the patient electronic health records), African-American, and male participants.

### 2.3 Empirical Analysis

To inform customization of financial incentives, there are several layers of information to be acquired. First, we established the statistical presence of preference heterogeneity towards incentive attributes among individuals (i.e., differences among individuals in their likes and dislikes) to provide support for the need for customization. We used a random parameter logit model assuming that preference parameters follow certain probability distributions (such as normal distribution or log-normal distribution) in order to allow them to differ across individuals. Second, we used a conditional logit model with interaction terms between attributes and demographic variables to investigate subgroup preference differences in the attribute ranking. The choice of demographic variables was based on the a priori identification of at-risk subgroups using evidence from the literature that includes gender, race, income, employment, education, and overweight/obesity status. All chosen demographic variables were included in the models simultaneously. Attributes that contribute more explanatory power to the model (and are

thus in some sense more ‘important’) would result in larger differences between the full and partial log-likelihood values. Third, rankings of levels within each attribute provide further insights into the design of subgroup-tailored incentive programs. We utilized the conditional log models estimated in the second step to conduct a counterfactual analysis. We systematically varied levels within a given attribute one at a time in a program condition (A) while holding other attributes at the same levels as the alternative program condition (B) and predicted the probabilities of each program being chosen relative to the opt-out choice. We used the differences of the predicted probabilities between A and B relative to the opt-out choice as the measure of the ‘marginal effect’ of a given attribute level on the program participation probability.

This last step involves probability prediction for individuals with specific demographic and health profiles such as an employed, Caucasian female with high education level, high income, and overweight. Those predicted probabilities are within-sample individual profile-specific predictions, which means that we need to have population-level compositions of those different at-risk subgroups to provide meaningful population-level reach information. Therefore, after predicting participation probabilities within our sample for each individual, we constructed sample weights to aggregate those individual probabilities to weighted average profile-level (e.g., national overall or subgroup levels) participation probability predictions to help inform incentive design selections. This approach follows Hensher et al. [33] and creates a synthetic population based on the adult US population, and predicts market shares (i.e., incentivized weight control program participation rates) in that synthetic population. To accomplish this goal, we used 2009–2010 National Health and Nutrition Examination Survey (NHANES) as a supplementary data set for the construction of the profile-specific weights. The NHANES is a nationally representative survey that contains demographic information and health outcomes of a national random sample of individuals. Each observation in the NHANES data set is assigned a sample weight that is designed as a measure of the number of people in the population represented by that sample person. We summed the individual sample weights provided in the NHANES data set across all NHANES individuals sharing the same individual profiles to generate profile-specific weights. To arrive at national level or more aggregated subgroup-level predictions, we predicted profile-specific participation probabilities and then used the profile weights to generate a weighted average of participation probability across profiles.

To facilitate a summary of the results, we used a ranking method to accommodate the nature of categorical attributes. For each individual profile, we ranked ordered levels within each attribute by each level’s ‘marginal effect’ on the participation probability (rank from 1 to 4 in order of importance or magnitudes of effect). Then, within each weighted

<sup>3</sup> Carilion Clinic is a tax-exempt health system headquartered in Roanoke, Virginia that serves all patients regardless of their ability to pay. It provides a comprehensive network of hospitals, primary and specialty physician practices and complementary services to about 1 million Virginians and West Virginians.

**Table 2** Comparison of study sample characteristics with Virginia state demographic characteristics

Variables	Sample (%)	Virginia (%)
<b>Race</b>		
White	49.0	71.3
Black	41.1	19.8
Native American	0.48	0.1
Asian	4.0	5.8
<b>Gender</b>		
Female	43.1	50.9
<b>Income</b>		
Household income \$10,000 or less	17.6	5.7
<b>Employment status</b>		
Percent unemployed	42.3	6.9
<b>Education</b>		
High school graduate or higher, percentage 25+	87.0	86.6
Bachelor’s degree or higher, percentage 25+	27.1	34.4
<b>Self-reported weight status</b>		
Underweight	0.1	2.0
Normal weight	7.2	34.4
Overweight	38.3	36.2
Obese	54.3	27.4

aggregate subgroup of interest, we calculated the proportion of that subgroup ranking each level of each attribute the same way. To confirm the statistical presence of preference heterogeneity towards incentive attributes in the sample, we estimated a random parameter logit model assuming all coefficients following normal distributions. To capture a log-linear relationship among reward magnitude and choice preference documented in the literature [34], we log-transformed the nonzero reward magnitude variable.

To investigate how many incentive options could maximize population reach and to what extent, we examined the program participation probability. We predicted the program participation probabilities under choice sets comparing single incentive options with multiple options. To maintain manageable option sets for our investigation, we fixed the only continuous attribute, reward amount, while we conducted the probability imputation. Then we changed the reward amount and fixed it while expanding the choice sets again. This way, we were able to see the distributions of participation probabilities change as the menu size changed under a given total amount of rewards.

### 3 Results

#### 3.1 Summary Statistics and Attribute Preference Heterogeneity

Summary statistics of key variables of our sample compared with Virginia’s census information is presented in Table 2.

The study sample contains a higher percentage of African-American, men, low-income, and unemployed individuals. These differences confirm the success of our strategy of oversampling those rarely studied groups (i.e., males, low socioeconomic status individuals). We also confirmed the existence of preference heterogeneity (Table 3). In summary, 60% of the coefficients (9 out of 15) have statistically significant standard deviations (SDs) signaling extensive preference heterogeneity within the sample. Furthermore, all incentive attributes have at least one level that demonstrated statistically significant SDs signaling the existence of preference heterogeneity in all attributes.

#### 3.2 Rankings of Attributes

This overall sample model (i.e., without interaction with demographic variables) shows that on average people mostly prefer a larger reward magnitude, to be paid weekly in the form of a debit card with contingency of attending weigh-ins and the program held in a community center. Next, we examined specific subgroup differences of their preferences towards each attribute. Figure 1 presents the attribute preference ranking results for the full sample and by subgroups that were classified by profile variables such as race, gender, income, and weight status. The rank ordering of attribute importance confirms the subgroup preference heterogeneity. The top two most important attributes in people’s program participation decision-making are dollar amount and program location. However, the relative ranking differs across subgroups. The obese subgroup and female subgroup exhibit

**Table 3** Random parameter logit model estimation results

Variables	Coefficients	
	Mean (SE)	Standard deviation (SE)
ASC	0.05 (0.113)	1.38*** (0.077)
Reward amount		
Log (magnitude)	0.36*** (0.042)	0.45*** (0.026)
Magnitude of \$0	0.18 (0.131)	0.86*** (0.130)
Program location [base: clinic]		
Workplace	− 0.24*** (0.048)	0.43*** (0.099)
Community center	0.17*** (0.048)	0.12 (0.116)
Church	− 0.18*** (0.048)	0.36*** (0.104)
Payment form [base: cash]		
Gym pass	− 0.13* (0.052)	0.50*** (0.115)
Medical debit card	− 0.18*** (0.049)	0.31*** (0.113)
Debit card	0.17*** (0.049)	0.06 (0.135)
Reward contingency [base: attendance]		
Weight	0.10 (0.053)	0.30* (0.120)
Compliance	− 0.06 (0.050)	0.18 (0.123)
Combined	−0.13** (0.048)	0.20 (0.114)
Payment frequency [base: once]		
Weekly	0.14* (0.054)	0.22 (0.128)
Monthly	− 0.02 (0.048)	0.10 (0.124)
Quarterly	0.05 (0.049)	0.29* (0.114)

ASC refers to alternative specific constant, which contains status quo choices

SE standard error

\* $p < 0.1$

\*\* $p < 0.05$

\*\*\* $p < 0.01$

the same preferences as the full sample. However, program location is deemed more important than dollar amount for subgroups of African Americans, low-income females, and males. Among those subgroups, the African-American subgroup exhibits the largest preference towards program location as compared with reward amount. This result is consistent with our experiences during focus groups, where African-American participants were particularly concerned with program location when discussing their choices (due to a number of reasons, such as convenience/accessibility, travel time, and privacy/stigma). Payment frequency is the least important attribute for most subgroups except for low-income females and females (whose least preferred attribute is reward condition).

### 3.3 Rankings of Levels Within Attributes

As shown in Table 3, the reward magnitude is positively correlated with individuals' willingness to participate in an incentivized weight control program holding everything else constant (i.e., mean effect of 0.36 with  $p < 0.01$ ). However, the effect's SD is sizable and statistically significant (SD

0.45;  $p < 0.01$ ). This means that there exist subgroups of individuals who dislike larger incentive amounts. We followed the profile participation probability generation process described above to predict participation probability with just the total reward amount varying from \$0–\$300 for the national aggregated and several subgroups of interest (i.e., profiles of obese individuals, African Americans, low-income females, males, and females). We present the prediction results in Fig. 2. Each panel shows the predicted participation probability distribution for each \$ amount for a specific aggregated population. The solid line shows the average participation probability. It is confirmed that the average participation responses to increasing reward amount are positive. Meanwhile, it is also clear that even the average responses show a sharp increase initially with a very small incentive amount. The dollar effects quickly reach their plateaus (i.e., beyond a small dollar amount threshold, the participation probabilities are no longer increasing at a high rate). For national level, it shows that the program will reach 50% participation with approximately \$40 total award for 6 months; however, \$100 total award (more than doubled the costs) is only projected to increase the participation rate

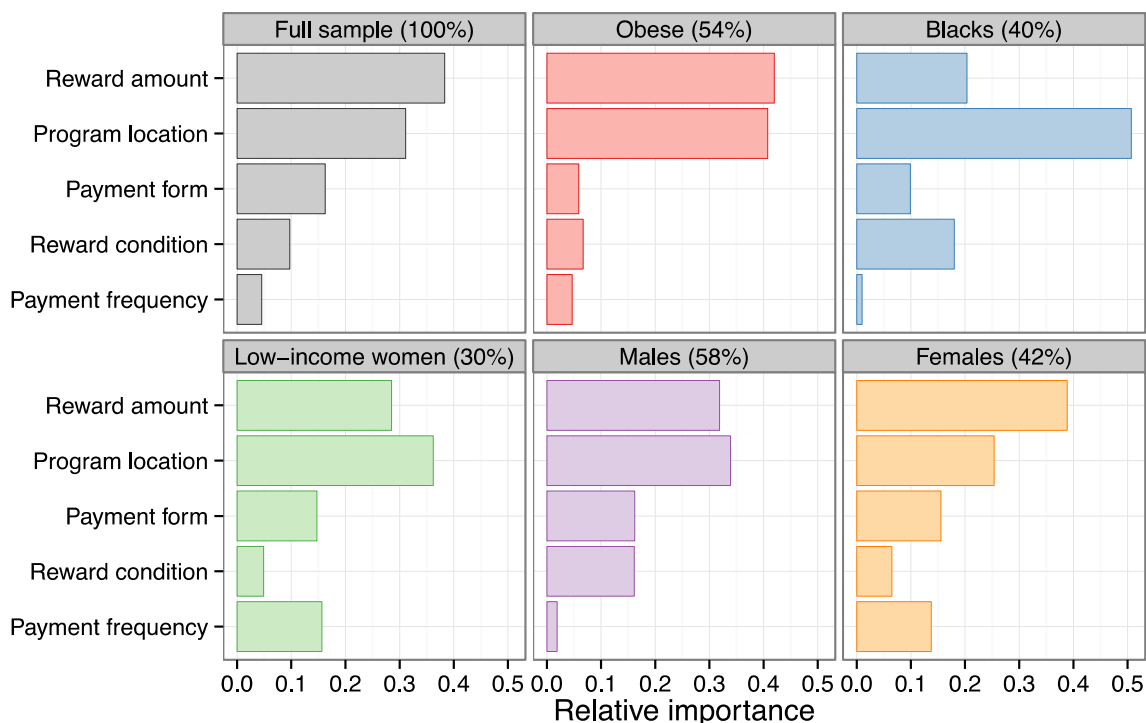


Fig. 1 Attributes’ relative importance (i.e., preference ranking) by populations of interest

by about 6%. This response trend is consistent across all populations examined.

We present the ranking results in Fig. 3. Each row of the figure corresponds to one incentive attribute. The first row is program location, the second row is payment form, the third row is reward contingency, and the fourth row is payment frequency. The four bars within the small panel correspond to the four populations of interest. A darker color patch signals a higher ranking (i.e., larger importance).

Using workplace as an example, the figure shows almost 50% of the national population would rank workplace as the most important program location level, while it was only ranked first for < 25% of the obese and African-American subgroups. For obese subgroups, clinic was the most preferred location choice. While for African-American and low-income women subgroups, community centers were the most preferred location. Program location is the one attribute that exhibits quite heterogeneous preferences across populations of interest.

Within the payment form attribute, the preference towards debit card was consistent across subpopulations except for African Americans. Almost 50% of African Americans responded to gym pass payment form the most, while about 35% responded to the debit card payment form the most. The least preferred payment form was medical debit card.

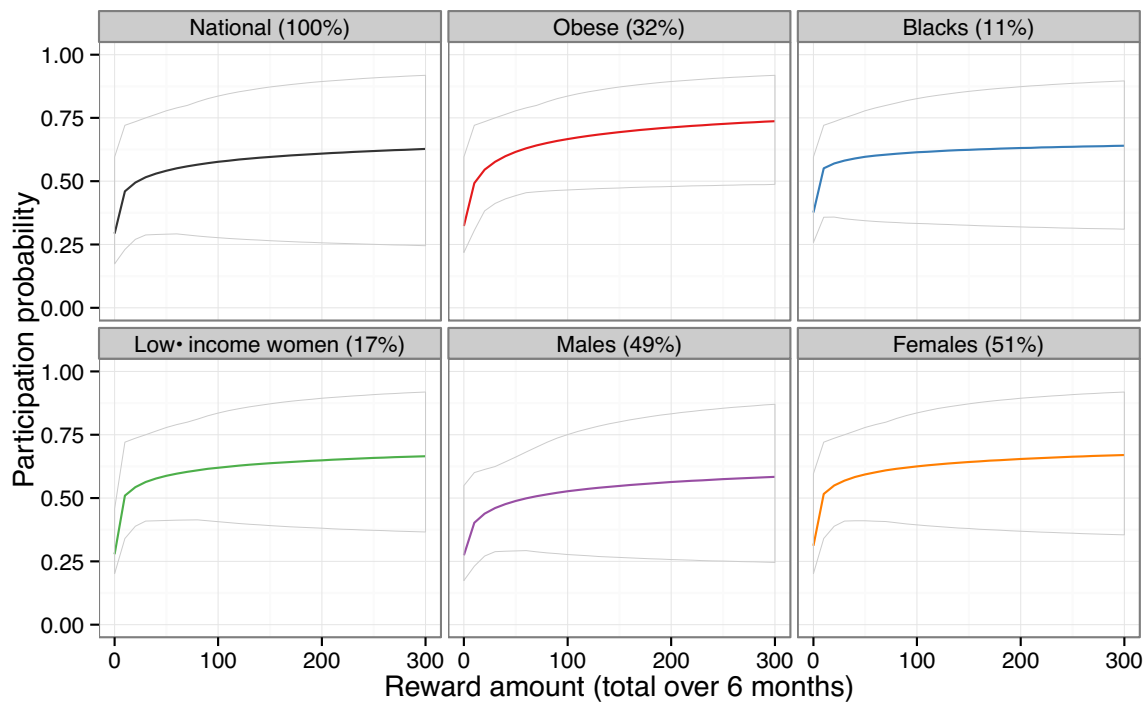
Similarly, most subgroups preferred rewards to be contingent on attending weigh-ins only (i.e., participants will

earn rewards as long as they show up to weigh-in). However, < 20% of African Americans preferred an attendance-only contingency. Instead, the African-American subgroup highly preferred incentives for weight loss (> 60% prefer weight contingency). The least preferred contingency condition is compliance. In terms of preference for payment frequency, it is quite consistent across all groups that weekly payment is the most desirable level with the obese subgroup showing the strongest preference towards this frequency level. The only notable difference is that the obese subgroup strongly disliked quarterly payments while the other subgroups rank quarterly payments the second-most preferred option. It is clear that payment at the end of the program is universally disliked.

### 3.4 Incentive Options

As shown in Table 3, all levels within program location and payment form show significant mean effects and most SDs are statistically significant.<sup>4</sup> The other two attributes have only one significant level and one significant SD within each. Reward contingency obtained a relatively larger statistical significance than payment frequency in terms of *p*-values.

<sup>4</sup> Among all the choices made, in about 24% the ‘opt-out’ option is selected.



**Fig. 2** The impacts of total reward amounts on program participation probability by populations of interest

However, it is of policy importance to know beyond the individual attribute and level significance in program reach to answer the following questions: (a) whether or not offering more than one incentive program option will achieve higher reach and if so to what extent; and (b) how many options are needed to provide the maximum population reach.

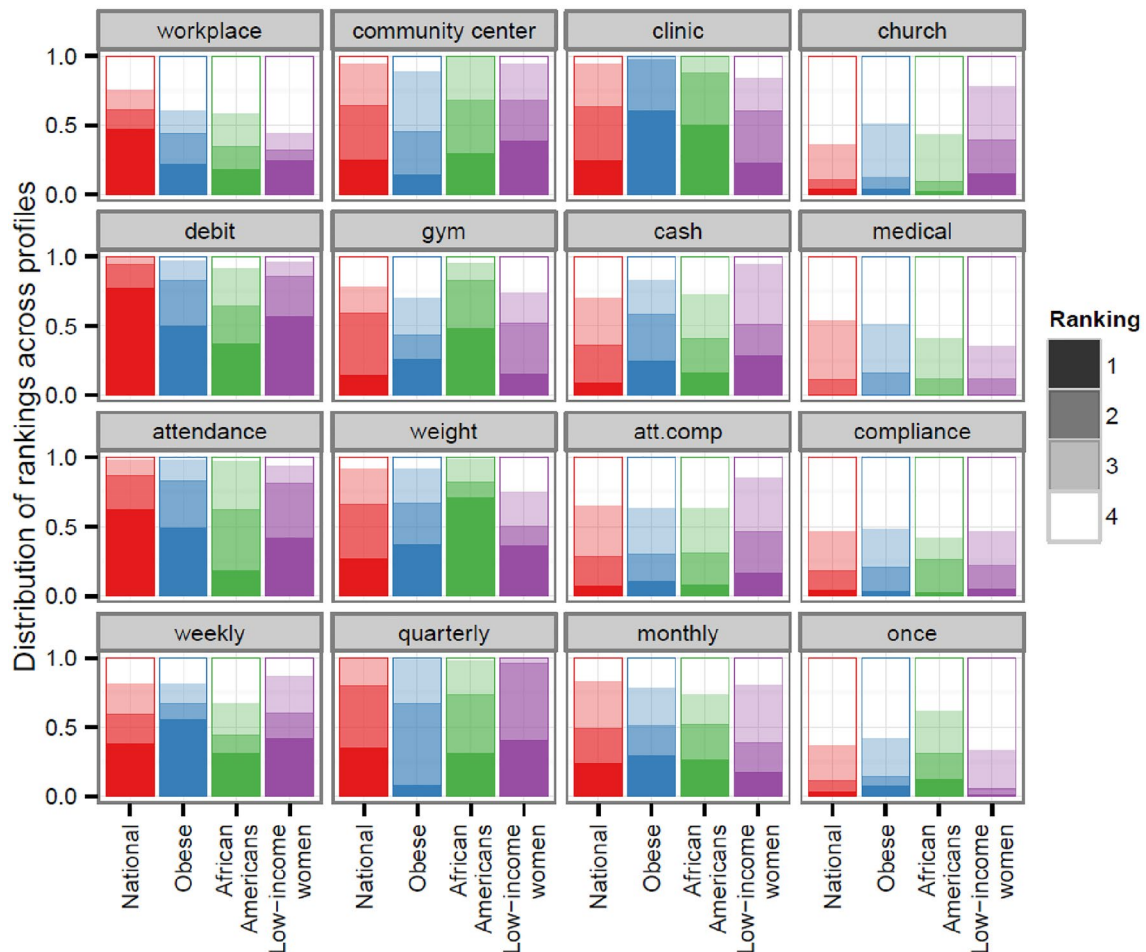
To answer those questions, we first ranked all the attributes (except for reward amount) by their relative importance and the dominance of heterogeneity informed by our model shown in Table 3. The order is program location, payment form, reward contingency, and then payment frequency. After ordering the attributes, we ordered levels within each attribute as well. We then examined the menu of options for each reward amount level.

For a given total reward amount, we constructed five option sets (menus) to examine the probability changes. The first menu contains only one incentive option and is designed with the first ranked level for each attribute (i.e., location of clinic, payment form of debit card, reward earned based on weigh-in attendance, reward paid at weekly frequency). The next menu offers two incentive designs which varied only by location levels (i.e., clinic vs community center), while other attributes maintain the same levels as the first menu. The third menu offers four incentive designs allowing both program location and payment form to take on two levels each, and so on until the fifth menu, which offers 32 incentive programs that allow all four attributes to take on two levels each.

We chose two levels of reward amounts to show the findings (shown in Fig. 4a [total reward amount is set at \$48] and Fig. 4b [total reward amount is set at \$576]). Both figures show the distributions of national-level participation probabilities with dots and lines representing average response levels for the national level and for those subgroups of interest (distributions for all subgroups show similar results, therefore we do not show them in one figure to maintain clarity). Both figures reveal a similar trend—as more incentive options are offered, the mean participation rate increases while the probability distributions shrink in range and in degree of dispersion. This result confirms that providing more incentive options for any given amount of rewards will have three benefits in population reach: first, it will increase program reach on average; second, it will reach more subpopulations and reduce program participation disparities; and third, it is easier for the program to recruit those hard-to-reach populations.

To elaborate on the third benefit, we examined the low end of the distributions. Offering two programs instead of one (two program locations vs one design) while setting total reward amount at \$48 would reach 54% of those hard-to-reach individuals instead of only 34% (a 59% increase in reach for those individuals whose willingness to participate is at the low end of the distribution) (Fig. 4a). The story remains the same when the total reward amount increases to \$576: offering two programs reaches 46% of those





**Fig. 3** Rankings by the impacts of attribute levels on program participation probability: program location, payment form, reward contingency, and payment frequency (by populations of interest)

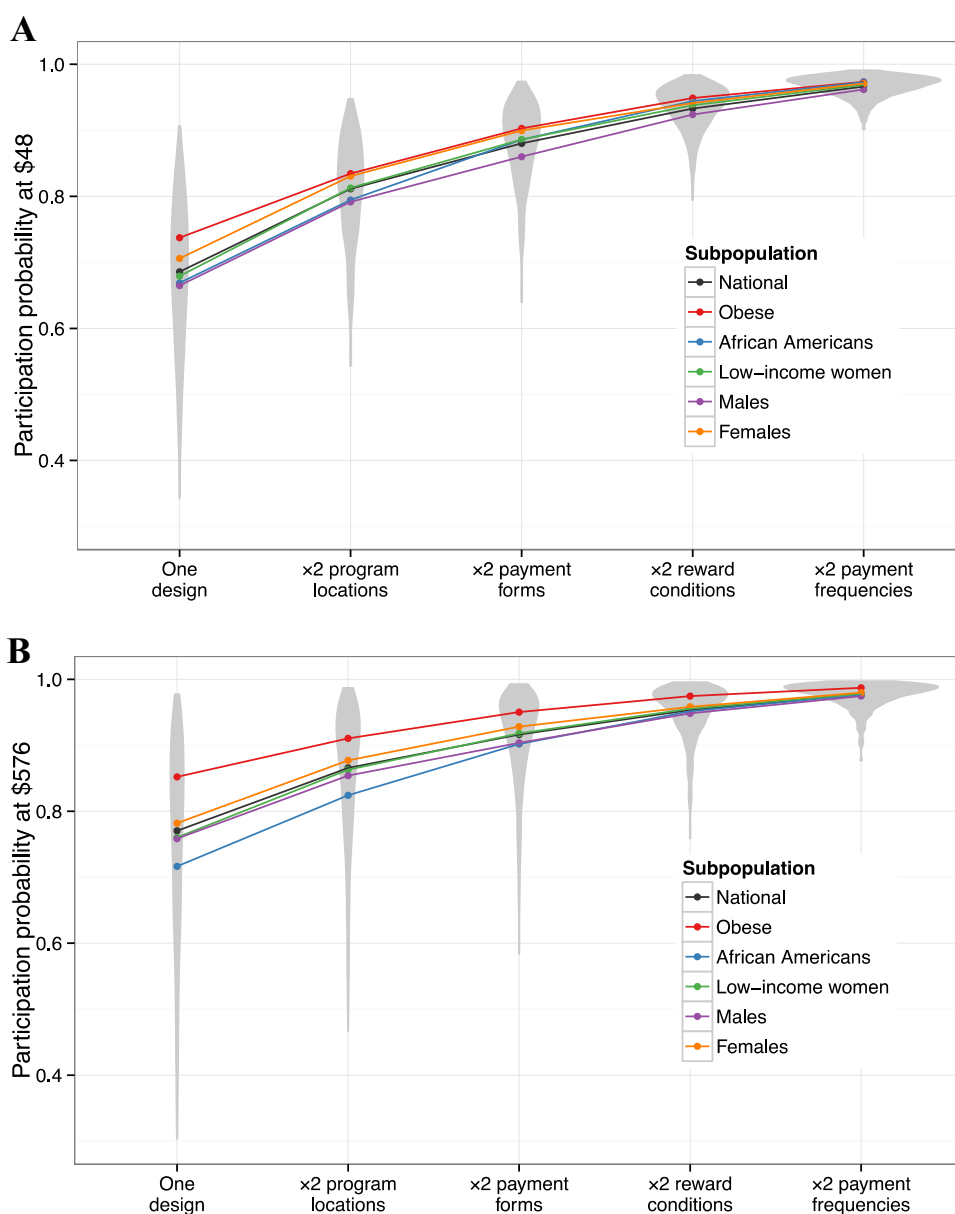
individuals while one program offering only reaches 16% of them (a 53% increase). Both figures confirm that the largest gain at return on choice sets is at offering two programs rather than one program. This conveys an important policy implication: if the goal is to increase population reach, offering two options will be the most cost-effective way.

Comparing Fig. 4a, b, it confirms the finding that ‘paying more’ does not necessarily do a better job at program reach. Offering a much higher reward amount does not gain more in reaching those hard-to-reach subgroups (a 53% increase at the \$576 level vs a 59% increase at the \$48 level). Furthermore, offering a larger reward results in larger disparities in program reach (i.e., the distribution range is larger than that with the smaller reward amount). For example, African Americans show a much lower mean participation rate than other subgroups when the dollar amount is set at \$576 as compared with \$48, signaling the existence of a much lower participation rate at the end of the quantiles for this subgroup.

### 3.5 Results Validation: A Case Study

To further examine the validity of our incentive design effect prediction and follow the recommendations of the literature [35, 36], we utilized a unique first-hand data set from a cluster randomized, controlled, worksite weight loss study [37, 38]. This study followed a two-group cluster randomized design. Worksites were randomly assigned into a comparison condition that did not include financial incentives for weight loss (i.e., Living My Way [LMW] condition) or an intervention condition that included a modest financial incentive tied to weight loss but intended to improve program reach (i.e., INCENTA condition; for details see [37]). The program lasted for 12 months. A total of 28 worksites were recruited and randomized into one of the two conditions. The study collected brief health information including self-reported weight and height and socio-economic and demographic information prior to the program recruitment, which provided a unique data set for analyzing program reach.

**Fig. 4 a** The impact of number of program options on the distribution of participation probability when total reward amount is set at \$2/week for 24 weeks by populations of interest. **b** The impact of number of program options on the distribution of participation probability when total reward amount is set at \$24/week for 24 weeks



The incentive design for the study can be summarized as the following: for the LMW group, there was \$0 in reward amount; there were monthly weigh-ins, and the program location was at the workplace; for the INCENTA group, the maximum reward was set at \$20/month; it was offered at the workplace; it was contingent on percentage of weight loss during weigh-ins, the reward was offered by check, and payment was quarterly. We used the baseline health survey data on all eligible employees ( $BMI \geq 24.5$ ) coupled with the closest attribute levels as the two conditions to predict the participation rate. The observed participation rates were 27% for control worksites and 36% for incentive worksites. Our model predicted that control worksites would have an approximately 25% participation rate (95% confidence

interval [CI] 15–39) and the incentive worksites would have an approximately 41% participate rate (95% CI 30–51). The prediction provides initial evidence of the validity of our model.

#### 4 Discussion and Policy Implications

There are sound scientific rationales for why incentives are effective and evidence from controlled clinical trials across a wide range of different applications supports their efficacy. Nevertheless, there are many important questions that remain to be answered regarding the use of incentives for health-related behavior changes, including

optimal incentive values, appropriate intervention durations for different problems and populations, and cost effectiveness [6, 18, 19, 39–41]. This gap in the literature is mainly due to the rareness of market data on the supply and demand of different variations of incentivized weight loss programs.

Our study presents a rare effort to address the preference for financial incentive designs to understand how population preferences may influence the reach of an incentivized weight control program. We considered greater numbers of attributes/levels and oversampled populations that were at risk for overweight/obesity and were understudied in the existing literature. To overcome the market data sparsity, we conducted an attribute-based choice experiment that enabled us to elicit and quantify target consumers' preference towards five types of attributes: reward amount, program location, payment form, reward condition, and payment frequency. As the results show, there are significant amounts of heterogeneous preferences across different populations of interest towards not only different attributes but also levels within each attribute. Even for the dollar amount, our common sense of 'paying more will never hurt' does not stand if the goal is to reach those hard-to-reach populations. Different incentive designs appeal to different types of individuals. Therefore, the natural conclusion is to offer a choice set of incentive options instead of a one-size-fits-all approach, with the goal of maximizing population reach and/or reaching those most needy subgroups.

Our random parameter logit model confirmed that all five attributes have a statistically significant impact on individuals' decision making in terms of incentivized weight control program participation. Among these attributes, reward amount, program location, and payment form were the most important attributes for all target subpopulations: the obese subgroup, African Americans, low-income women, females, and males. The order of importance among them varies by subpopulation. Our results reveal that paying more does not help with the goal of reaching some subgroups. For example, program location is more important in decision making for African Americans, low-income females, and males compared with reward amount. African Americans specifically showed a negative response to larger reward amounts; paying them more may actually push some of them away from the program instead of attracting them in. Although decreasing willingness to participate with increasing reward amount seems counterintuitive, the phenomenon matches focus group observations and previous literature [42]. Some individuals in the focus groups reacted negatively to the presence of a financial reward, expressing the sentiment that individuals should lose weight for their health, not for financial gain. Indeed, financial incentives in behavioral health interventions have already raised a degree of ethical controversy [43–52]. Our findings suggest

that low levels of reward amount are appropriate and more cost effective at closing the program reach disparity gap because they minimize the impact from negative responses and also gain most of the participation responses from positive responses.

Our results further show that preference heterogeneity exists even within each attribute. For example, among program location levels, workplace was the most preferred location for the national population while it was less preferred for some subgroups such as the obese subgroup (where a clinic was the most preferred location) and the African-American subgroup (a community center was most preferred). Similarly, debit card and gym pass were the two top ranked levels within payment form, and the order varied by subgroup. There was relatively more consistency in the preference for levels within reward contingency and payment frequency where contingency on weigh-in attendance and weekly payment schedule were the most preferred options. There still exist vast differences among one or two subgroups regarding those. For example, African Americans mostly prefer contingency on weight loss outcomes directly. There is universal dislike towards medical card and payment at the end of the program.

Furthermore, we used our models to conduct a systematic counterfactual analysis through predicting participation probability weighted by a NHANES national sample weight for meaningful external prediction results. This exercise revealed several policy-relevant insights: offering more than one incentive option will not only gain the program's population reach but will also increase the participation rate among those hard-to-reach subgroups; offering a menu of two programs instead of one will gain the most in participation boost; offering a larger reward amount will work against the goal of reaching those hard-to-reach subgroups.

Several limitations should be kept in mind when interpreting our findings. First, our sample is recruited from one local electronic medical record, which limits the generalizability of our findings. The synthetic population construction using NHANES sample weights is helpful to mitigate this sampling limitation but cannot address it fully since it only controls for observable constructs. If the local unobserved differences (e.g., norms) drive preference heterogeneity, our findings cannot be generalized to other contexts. Second, the hypothetical nature of the discrete choice experiment is subject to hypothetical bias. However, we followed the best practices of the discrete choice experiment literature to mitigate this bias through using 'the cheap talk' [53]. 'The cheap talk' is a text script presented to the respondents before they see the choice questions, which highlights the importance of the real-world impact of their answers and alerts them to possible bias. Furthermore, our results focus on examining relative preference differences between subgroups; therefore,

as long as the bias is homogeneous across subgroups the relative differences findings would cancel out or minimize the bias.

Some future work is needed to further investigate causality behind the preference towards incentive designs we found; for example, investigating causes for the negative response to reward amount to design recruitment protocols to ameliorate those unintended spillovers. Previous cases of repugnance have been successfully treated with in-kind rather than monetary transactions [54], but given the overwhelming preference for payment with debit card we saw in our study, a less desirable payment form may reduce participation probability more than it increases by not being monetary. Future work could also investigate program location more closely, since it is the second-most important attribute and is also the one that displays a great deal of heterogeneity. Program location itself could be decomposed into constituent attributes, which may provide a more coherent picture of location preferences. Understanding why some individuals prefer some locations over others can suggest locations not previously studied, new locations that could be built, or predict which existing location would be most suitable given a target population and geographic area.

Despite considering more incentive components than previous studies, there are still more ways in which financial incentives can vary. Evidence suggests that deposit contracts lower participation rates [24], and future work could ascertain this result as well as check for individual heterogeneity. Escalating payments, in which later payments are larger than earlier payments, may increase effectiveness [55] or self-weighting [56], but their effect on participation is unknown. Adding uncertainty to the incentive design may

also increase effectiveness [26, 57], but again, the effect on participation is unknown. Previous studies suggest that group-based incentives may be more powerful than individual incentives [19, 58], but the effect on participation and the role of individual heterogeneity is unknown. Exploiting group dynamics may be an inexpensive way to increase reach and effectiveness.

## 5 Conclusion

Effective weight control programs will achieve high population health impact if they can reach those needy subpopulations. Incentivized weight control programs can increase program participation but the incentive schemes should be tailored to subgroups that exhibits different incentive preference to achieve its population health impact. The study detangles the preference for incentive attributes with the goal of improving implementation effectiveness. We confirmed the existence of preference heterogeneity and the importance of targeted incentive design in maximizing population reach and closing health disparities. Our findings also show that offering a menu of incentive designs (more than one design option) for participants to choose from would increase program participation rate more than offering higher reward amounts. This indicates that offering more choices will be more cost-effective at increasing program reach.

## Appendix

See Figs. 5, 6 and 7.





Fig. 6 The financial incentive construct description in the survey

**1 Please consider the following two weight control programs.**

	<b>Program A</b>	<b>Program B</b>
<b>Program location</b>	Clinic	Church
<b>Weekly weight control goal</b>	Attending weekly weight checks	Turning in records of diet and exercise
<b>Weekly reward available</b>	\$16	\$4
<b>Payment frequency</b>	Once at end of program	Weekly
<b>Payment form</b>	Health debit card for doctor's visits, prescriptions, and other medical expenses	Debit card
<b>Total reward available in program</b>	\$16 × 24 wks = \$384	\$4 × 24 wks = \$96

**Which weight control program would you choose?**  
*Please check one box.*

<sub>1</sub> Program A

<sub>2</sub> Program B

<sub>3</sub> I would not choose either program.

114 10

**Fig. 7** Example of one program choice set presented in the survey. Each survey participants will see four sets of choices comprising different combinations of program A and B

## Declarations

**Funding** Research reported in this publication was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under award number 1R21HL097308-01A2. The content is solely the responsibility of the authors and does not necessarily represent the official views of the sponsoring agencies.

**Conflicts of Interest** Wen You, Yuan Yuan, Kevin Boyle, Tzeyu Michaud, Chris Parmeter, Richard Seidel, and Paul Estabrooks declare that they have no conflict of interest.

**Ethics approval** The research was approved by the Virginia Tech Institutional Review Board (FWA00000572) and by the Carilion Clinic Institutional Review Board.

**Consent to participate** Focus group participants consented through written informed consent and survey participants provided oral consent over the phone during the recruitment process.

**Consent for publication** Not applicable since all results will be anonymous and at an aggregated level.

**Availability of data and material** The de-identified data will be made available upon request.

**Code availability** The software used are R and Stata. The codes can be made available upon request.

**Authors' contributions** WY, KB and PE were responsible for the conceptualization, methodology, and execution. WY and PE were responsible for the write up of the first draft of the paper. WY, KB, YY, RS, and PE were responsible for the survey design, recruitment, and survey administration. YY, WY, KB, CP, and PE analyzed and interpreted the data. All authors participated in the manuscript writing.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.


## References

- Anderson LM, et al. The effectiveness of worksite nutrition and physical activity interventions for controlling employee overweight and obesity a systematic review. *Am J Prev Med.* 2009;37(4):340–57.
- Hillier-Brown FC, et al. A systematic review of the effectiveness of individual, community and societal-level interventions at reducing socio-economic inequalities in obesity among adults. *Int J Obes (Lond).* 2014;38(12):1483–90.
- Haughton CF, et al. Racial/ethnic representation in lifestyle weight loss intervention studies in the United States: a systematic review. *Prev Med Rep.* 2018;9:131–7.
- Coupe N, Cotterill S, Peters S. Tailoring lifestyle interventions to low socio-economic populations: a qualitative study. *BMC Public Health.* 2018;18(1):967.
- Hillier-Brown FC, et al. A systematic review of the effectiveness of individual, community and societal level interventions at reducing socioeconomic inequalities in obesity amongst children. *BMC Public Health.* 2014;14:834.
- Finkelstein EA, Bilger M, Baid D. Effectiveness and cost-effectiveness of incentives as a tool for prevention of non-communicable diseases: a systematic review. *Soc Sci Med.* 2019;232:340–50.
- Gong Y, et al. Financial incentives for objectively-measured physical activity or weight loss in adults with chronic health conditions: a meta-analysis. *PLoS ONE.* 2018;13(9): e0203939.
- Mitchell MS, et al. Financial incentives for physical activity in adults: systematic review and meta-analysis. *Br J Sports Med.* 2019;54:1259–68.
- Ananthapavan J, Peterson A, Sacks G. Paying people to lose weight: the effectiveness of financial incentives provided by health insurers for the prevention and management of overweight and obesity—a systematic review. *Obes Rev.* 2018;19(5):605–13.
- Tambor M, et al. Financial incentives for a healthy life style and disease prevention among older people: a systematic literature review. *BMC Health Serv Res.* 2016;16(5):426.
- Paul-Ebhohimhen V, Avenell A. Systematic review of the use of financial incentives in treatments for obesity and overweight. *Obes Rev.* 2008;9(4):355–67.
- Sykes-Muskett BJ, et al. The utility of monetary contingency contracts for weight loss: a systematic review and meta-analysis. *Health Psychol Rev.* 2015;9(4):434–51.
- Kullgren JT, et al. A scoping review of behavioral economic interventions for prevention and treatment of type 2 diabetes mellitus. *Curr Diab Rep.* 2017;17(9):73.
- Kane RL, et al. A structured review of the effect of economic incentives on consumers' preventive behavior. *Am J Prev Med.* 2004;27(4):327–52.
- Paloyo AR, et al. The causal link between financial incentives and weight loss: an evidence-based survey of the literature. *J Econ Surv.* 2014;28(3):401–20.
- World Health Organization. Obesity: preventing and managing the global epidemic. World Health Organization; 2000.
- Volpp KG, et al. Redesigning employee health incentives—lessons from behavioral economics. *N Engl J Med.* 2011;365(5):388–90.
- Higgins ST, et al. Incentives and health: an introduction. *Prev Med.* 2012;55:S2.
- Jeffery RW. Financial incentives and weight control. *Prev Med.* 2012;55:S61–7.
- Cawley J. Incentives for wellness: tackling issues of effectiveness, biology, scope, and cost shifting. *J Policy Anal Manag.* 2014;33(3):826–32.
- Mattke S, et al. Workplace wellness programs study. *Rand Health Q.* 2013;3(2):7.
- Gneezy U, Rustichini A. Pay enough or don't pay at all. *Q J Econ.* 2000;115(3):791–810.
- Frey BS, Jegen R. Motivation crowding theory. *J Econ Surv.* 2001;15(5):589–611.
- Jeffery RW, Hellerstedt WL, Schmid TL. Correspondence programs for smoking cessation and weight control: a comparison of two strategies in the Minnesota Heart Health Program. *Health Psychol.* 1990;9(5):585.
- Butsch WS, et al. Effects of a reimbursement incentive on enrollment in a weight control program. *Obesity.* 2007;15(11):2733–8.



26. Cawley J, Price JA. Outcomes in a program that offers financial rewards for weight loss. National Bureau of Economic Research; 2009.
27. Forster JL, et al. A work-site weight control program using financial incentives collected through payroll deduction. *J Occup Med*. 1985;27(11):804–8.
28. Roux L, et al. Valuing the benefits of weight loss programs: an application of the discrete choice experiment. *Obes Res*. 2004;12(8):1342–51.
29. Owen K, et al. Individual preferences for diet and exercise programmes: changes over a lifestyle intervention and their link with outcomes. *Public Health Nutr*. 2010;13(2):245–52.
30. Webb EJD, et al. Measuring commissioners' willingness-to-pay for community based childhood obesity prevention programmes using a discrete choice experiment. *BMC Public Health*. 2020;20(1):1535.
31. Coast J, et al. Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations. *Health Econ*. 2012;21(6):730–41.
32. Johnson FR, et al. Experimental design for stated-choice studies. In: *Valuing environmental amenities using stated choice studies*. Springer; 2006. p. 159–202.
33. Hensher DA, Rose JM, Greene WH. *Applied choice analysis: a primer*. Cambridge University Press; 2005.
34. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making. *Pharmacoeconomics*. 2008;26(8):661–77.
35. Quaife M, et al. How well do discrete choice experiments predict health choices? A systematic review and meta-analysis of external validity. *Eur J Health Econ*. 2018;19(8):1053–66.
36. Mohammadi T, et al. Testing the external validity of a discrete choice experiment method: an application to latent tuberculosis infection treatment. *Value Health*. 2017;20(7):969–75.
37. Almeida FA, et al. Effectiveness of a worksite-based weight loss randomized controlled trial: the worksite study. *Obesity (Silver Spring)*. 2015;23(4):737–45.
38. You W, et al. Who participates in internet-based worksite weight loss programs? *BMC Public Health*. 2011;11(1):709.
39. John LK, Loewenstein G, Volpp KG. Empirical observations on longer-term use of incentives for weight loss. *Prev Med*. 2012;55:S68–74.
40. Parkinson B, et al. Designing and using incentives to support recruitment and retention in clinical trials: a scoping review and a checklist for design. *Trials*. 2019;20(1):624.
41. Thirumurthy H, Asch DA, Volpp KG. The uncertain effect of financial incentives to improve health behaviors. *JAMA*. 2019;321(15):1451–2.
42. Wanders JO, et al. The effect of out-of-pocket costs and financial rewards in a discrete choice experiment: an application to lifestyle programs. *BMC Public Health*. 2014;14(1):870.
43. Halpern SD, Madison KM, Volpp KG. Patients as mercenaries? The ethics of using financial incentives in the war on unhealthy behaviors. *Circ Cardiovasc Qual Outcomes*. 2009;2(5):514–6.
44. Pearson SD, Lieber SR. Financial penalties for the unhealthy? Ethical guidelines for holding employees responsible for their health. *Health Aff*. 2009;28(3):845–52.
45. Schmidt H, Voigt K, Wikler D. Carrots, sticks, and health care reform—problems with wellness incentives. *N Engl J Med*. 2010;362(2): e3.
46. Schmidt H. Wellness incentives, equity, and the 5 groups problem. *Am J Public Health*. 2012;102(1):49–54.
47. Schmidt H, Asch DA, Halpern SD. Fairness and wellness incentives: what is the relevance of the process-outcome distinction? *Prev Med*. 2012;55:S118–23.
48. Ashcroft RE. Personal financial incentives in health promotion: where do they fit in an ethic of autonomy? *Health Expect*. 2011;14(2):191–200.
49. Madison KM, Volpp KG, Halpern SD. The law, policy, and ethics of employers' use of financial incentives to improve health. *J Law Med Ethics*. 2011;39(3):450–68.
50. Reisinger HS, et al. "All the money in the world..." patient perspectives regarding the influence of financial incentives. *Health Serv Res*. 2011;46(6pt1):1986–2004.
51. Klein EP. Patient health incentives: ethical challenges and frameworks. *Int J Behav Med*. 2014;21(6):995–1004.
52. Lunze K, Paasche-Orlow MK. Financial incentives for healthy behavior: ethical safeguards for behavioral economics. *Am J Prev Med*. 2013;44(6):659–65.
53. List J, Gallet C. What experimental protocol influence disparities between actual and hypothetical stated values? *Environ Resour Econ*. 2001;20:13.
54. Roth AE. Repugnance as a constraint on markets. *J Econ Perspect*. 2007;21(3):37–58.
55. Stitzer M, Petry N. Contingency management for treatment of substance abuse. *Annu Rev Clin Psychol*. 2006;2:411–34.
56. Yancy WS, et al. Effect of escalating financial incentive rewards on maintenance of weight loss: a randomized clinical trial. *JAMA Netw Open*. 2019;2(11):e1914393–e1914393.
57. Volpp KG, et al. Financial incentive-based approaches for weight loss: a randomized trial. *JAMA*. 2008;300(22):2631–7.
58. Kullgren JT, et al. Individual vs. group-based incentives for weight loss: a randomized, controlled trial. *Ann Intern Med*. 2013;158(7):505.

## Authors and Affiliations

Wen You<sup>1</sup>  · Yuan Yuan<sup>2</sup> · Kevin J. Boyle<sup>3</sup> · Tzeyu L. Michaud<sup>4</sup> · Chris Parmeter<sup>5</sup> · Richard W. Seidel<sup>6</sup> · Paul A. Estabrooks<sup>4</sup>

Yuan Yuan

y.yuan@vt.edu

Kevin J. Boyle

kjboyle@vt.edu

Tzeyu L. Michaud

tzeyu.michaud@unmc.edu

Chris Parmeter

c.parmeter@miami.edu

Richard W. Seidel

rwseidel@carilionclinic.org

Paul A. Estabrooks

paul.estabrooks@unmc.edu

<sup>1</sup> Department of Public Health Science, School of Medicine, University of Virginia, University of Virginia Cancer Center, Charlottesville, VA, USA

<sup>2</sup> Convoy Inc., Seattle, WA, USA

<sup>3</sup> Department of Agricultural and Applied Economics, Willis Blackwood Real Estate, Virginia Tech, Blacksburg, VA, USA

<sup>4</sup> Department of Health Promotion, University of Nebraska Medical Center, Omaha, NE, USA

<sup>5</sup> Department of Economics, University of Miami, Miami, FL, USA

<sup>6</sup> Department of Psychiatry and Behavioral Medicine, Virginia Tech Carilion School of Medicine, Roanoke, VA, USA