Abstract

There is an increased interest in the use of personalized medicine approaches in the prevention or treatment of obesity, however, few studies have used these approaches to identify individual differences in treatment effects. The current study demonstrates the use of the predicted individual treatment effects (PITE) framework to test for individual differences in the effects of the ACTION-PAC intervention, which targeted the treatment and prevention of obesity in a high school setting. We show how methods for personalized medicine can be used to test for significant individual differences in responses to an intervention and we discuss the potential and limitations of these methods. In our example, 25% of students in the preventive intervention, were predicted to have their BMI z-score reduced by 0.39 or greater, while at other end of the spectrum, 25% were predicted to have their BMI z-score increased by 0.09 or more. In this paper, we demonstrate and discuss the process of using methods for personalized medicine with interventions targeting adiposity and discuss the lessons learned from this application. Ultimately, these methods have the potential to be useful for clinicians and clients in choosing between treatment options, however they are limited in their ability to help researchers understand the mechanisms underlying these predictions.

The prevalence estimates of obesity in adolescents in the United States indicate that 1 in 5 youth have a body mass index (BMI) classified as obese (≥95th percentile for their age- and sex) (Hales et al., 2018). Childhood obesity can lead to long-term health problems including cardiovascular disease, osteoarthritis, diabetes, and certain types of cancer (Chatelan et al., 2019; Preston et al., 2018). Although many interventions have been developed to reduce adiposity or prevent obesity in youth (O'Connor et al., 2017; Wang et al., 2015; Wilson et al., 2015), the average effect size for most is quite small, often 0.20 or less (O'Connor et al., 2017). Importantly, across various types of obesity interventions (lifestyle, pharmacological, and surgical), there is considerable heterogeneity of response (Heymsfield et al., 2018; Kelly et al., 2018; MacLean et al., 2018). This is unsurprising due to a complex interplay between biological, environmental and socio-behavioral etiologies of obesity. Root causes and effective treatments are therefore likely to differ between individuals (Baranowski et al., 2019; Beets et al., 2019; Hoelscher et al., 2013). Because the etiology of obesity differs and interventions target obesity through different mechanisms (e.g., diet, physical activity, motivation), it is reasonable that the optimal intervention is likely to vary for different individuals (Baranowski et al., 2019; Beets et al., 2019; Hoelscher et al., 2013; MacLean et al., 2018; Yanovski & Yanovski, 2018).

There has recently been a push for the use of personalized medicine to help choose the treatment most likely to work for an individual (Heymsfield et al., 2018; Ingelsson & McCarthy, 2018; Kelly et al., 2018; MacLean et al., 2018; Yanovski & Yanovski, 2018). The current paper utilizes a biopsychosocial framework and a specialized analytic strategy for estimating predicted individual treatment effects (PITEs) (Ballarini et al., 2018; Lamont et al., 2018) using data from the ACTION-PAC trial (Vallabhan et al., 2017). We aim to demonstrate the potential for using new methods for personalized medicine to assess individual differences in response to an

intervention targeting adiposity and clarify the limitations of these methods. We use PITEs to test for evidence of heterogeneity in the effects of both treatments, describe individual differences, and identify variables that contribute the most to these differences. We then discuss the lessons learned from this application and implications for future directions needed to increase the probability that methods for personalized medicine will impact clinical practice.

The effect of interventions for the prevention and treatment of obesity

In the last 20 years, there have been over 100 randomized trials of interventions targeting the prevention of obesity or reductions in adiposity in children or adolescents (Wang et al., 2015). Meta-analyses of these trials indicate that effective intervention models tend to utilize intensive, multi-component approaches (O'Connor et al., 2017; Wang et al., 2015; Irvin & Kaplan, 2016). However, even the most intensive treatment trials had small effect sizes (Cohen's D=0.22-0.34). Less intensive trials typically show very small average effect sizes <0.10 (O'Connor et al., 2017). Similar effects are observed in preventive interventions, with most producing effect sizes <0.10 (Wang et al., 2015). That intensive, multi-component interventions are required to achieve even moderate effects suggests the need to improve efficacy of treatments (Heymsfield et al., 2018; Kelly et al., 2018; MacLean et al., 2018) and that individuals may respond differently to different treatment approaches (Severin et al., 2019). However, efforts to identify predictors of individual differences in treatment effect have found that the only reliable predictors are adherence to treatment protocols (Lemstra et al., 2016; Severin et al., 2019) and initial response to treatment (e.g., weight lost during the first weeks) (Heymsfield et al., 2018; Yanovski & Yanovski, 2018).

Personalized medicine for the treatment of obesity

In an effort to improve efficacy of interventions targeting obesity, there is a focus on the

potential role of personalized medicine (Estampador & Franks, 2018; Frühbeck et al., 2018; Ingelsson & McCarthy, 2018; Yanovski & Yanovski, 2018). The premise of 'personalized medicine' (Kent et al., 2018; Webb et al., 2020) is that by predicting individual differences in treatment effects, these differences can be used to select the most effective treatment for each individual. Thus, our working definition of personalized medicine is "the use of individual level data to choose the treatment most likely to be successful for an individual." Personalized medicine aims to improve average outcomes for the entire population through individual targeting (Kranzler & McKay, 2012; Smith et al., 2013).

To date, most of the publications on personalized medicine for obesity have been conceptual (Estampador & Franks, 2018; Ingelsson & McCarthy, 2018), with authors making the argument that personalized approaches will be needed to increase treatment efficacy (Frühbeck et al., 2018; Yanovski & Yanovski, 2018). Recent research with adolescent samples has quantified a high degree of heterogeneity in response to treatment and argued for personalized approaches (Ryder et al., 2019). We know of only one paper which applies new predictive personalized medicine methods to obesity treatment, focusing on identifying differential responders for adults with osteoarthritis (Xiaotong et al., 2020). Thus, the argument for personalized medicine in the treatment and prevention of obesity remains largely conceptual with those in the field arguing for a holistic, biopsychosocial approach (Frühbeck et al., 2018; Ingelsson & McCarthy, 2018; Yanovski & Yanovski, 2018). This suggests that the methods used to study individual responses to obesity prevention and treatment approaches should be able to detect heterogeneity in treatment effects due to complex biological, psychological, and social factors and examine higher-order and non-linear interactions.

New methods for personalized medicine

We distinguish analytic approaches to personalized medicine such as using interactions to identify moderators of treatment effects, from new *methods for personalized medicine*. Established methods are designed to test specific hypotheses about individual differences. The new methods focus on individual predictions, rather than hypothesis testing. These methods utilize predictive algorithms, often using machine learning approaches, and are typically based on the potential outcomes framework (Holland, 1986; Rubin, 2005). Established approaches for testing *a priori* hypothesized interactions between treatment and variables expected to predict differential treatment response are effective when a small number of key variables are expected to moderate treatment effects. In contrast, personalized medicine methods work when heterogeneity in treatment effects is more complex or when there are expected to be a large number of moderators (Green & Kern, 2012; Henderson et al., 2017; Imai & Strauss, 2011; Poulson, 2011). A strength of these methods is that they are designed to directly address the goal of improving treatment decisions by providing clinicians and clients with practically relevant predictions to help guide selection of the most appropriate intervention for a particular patient. A major limitation of these methods is that they often have limited ability to help researchers in the area understand the underlying mechanisms behind why the interventions work or fail to work. Methods to test mediation and moderation are among the approaches designed for testing mechanisms.

The new methods work best when theory, previous literature and clinical knowledge is used in selecting predictors and for guiding the choice of predictive model (Hoogland et al., 2021). However, these methods are not designed to isolate moderating effects of particular variables. They instead focus on the larger picture and constellation of interactions between many relevant variables. In reference to this constellation of interactions, Breiman (2001) argued

that a “forest of trees is impenetrable as far as simple interpretations of mechanisms go.” Machine learning methods, such as Random Forests or BART, facilitate this approach to personalized medicine by picking up on the subtle relationships between predictors, each of which may have only a small impact. To provide this predictive accuracy regarding optimal treatment options for a particular patient, these methods eschew focusing on particular predictors in favor of assessing many potentially relevant variables simultaneously. Even though the mechanisms may not be understood, under conditions which are typically met in randomized trials, the results of the PITE framework and other similar methods can be interpreted as causal effects (Hoogland et al., 2021). However, a disadvantage of these methods is that, because they do not provide tests of mechanisms, options for diagnosing failure to find individual differences or a failure to replicate results are limited.

The goal of this paper is to demonstrate the use and limitations of one of these methods, the PITE framework (Ballarini et al., 2018; Kuhlemeier et al., 2021; Lamont et al., 2018), for examining the utility of personalized medicine methods in obesity research. PITE is especially relevant to obesity treatment because of its focus on individual predictions, rather than identifying subgroups, and its flexibility in which predictive method can be used. Many existing methods focus on identification of subgroups who respond differently to treatment. Estimating individual effects is more appropriate when many different moderators jointly determine treatment response, which is what is expected in interventions targeting complex conditions like obesity.

Using a potential outcome framework (Angrist et al., 1996; Holland, 1986; Rubin, 2005), the causal effect of a treatment for client *i* is defined as the difference between their potential outcome under both treatment conditions, which is the outcome that they would obtain if

assigned to treatment (Y_i^t) and the outcome they would obtain under control (Y_i^c). Thus, the causal effect of treatment for an individual is defined as:

$$Y_i^t - Y_i^c \quad (1)$$

Because the client cannot receive both conditions at the same time, the causal effect for any client is never observable. Instead we use a predictive algorithm with baseline covariates, *X*, and observed outcomes from those in the original randomized trial to obtain predictions of the potential outcome for each individual as a function of their values on the baseline covariates (Lamont et al., 2018):

$$\text{PITE}_i = \hat{Y}_i^t - \hat{Y}_i^c = f_t(x_i) - f_c(x_i) \quad (2)$$

Where f_t is a function relating covariates for individual *i* to a predicted outcome under treatment and *fc* is this function under control. Thus, the PITE for a particular individual is the predicted effect of the intervention for them given their covariates and the particular function used to make the estimates. The algorithm for computing PITEs is:

1) Fit a predictive model or algorithm to the outcome using those in the control condition.
2) Fit another predictive model/algorithm to the outcome for the treatment condition.
3) For any individual, *i*, with observed covariates *x*, predictions of the outcome under treatment and control are computed by applying the predictive methods in 1 and 2.
4) The PITE for individual *i* is the difference between their predicted outcome under treatment and control.

One strength of the PITE framework is that it can be used with any predictive model or algorithm that allows individual-level outcome prediction. This approach focuses on including information from many variables that, together, can predict individual differences. Thus, PITE is equivalent to simultaneously testing many moderators. When BART (other machine learning

predictive methods also have the same property) is used, it tests higher-order and non-linear interactions. Analyses using PITE can produce predictions of intervention effects for any individual for whom covariates can be measured. As such, PITE can generate predictions for individuals not originally in the randomized clinical trial.

Current Study

This study aims to demonstrate the use of methods for personalized medicine for obesity prevention and treatment using data from the ACTION-PAC cluster-randomized trial (ClinicalTrials.gov ID: NCT02502383), which tested two related interventions, a very low intensity preventive intervention for students with BMI < 85th percentile and a low-intensity (approximately 6 hours of contact time over 2 years) intervention for those with BMI > 85th percentile administered by school-based health center (SBHC) providers. We estimate individual differences in the effect of the ACTION-PAC intervention on endpoint BMI z-scores two years after the start of the intervention. The effects of the preventive and intensive interventions were estimated separately because these interventions differed substantively including in their level of intensity.

Our analysis first examines whether there is evidence for individual differences in the effects of each treatment. Where individual differences are found, we describe the range of treatment effects estimated for that treatment. Finally, we conduct a variable importance analysis to identify variables which may be driving the observed differences. Our goal in this study is to show the potential use of methods for personalized medicine in obesity research, to demonstrate what these methods can contribute, and to discuss their limitations. One limitation of this study is that while the study was cluster-randomized, as of yet neither PITE, nor to our knowledge any other methods for personalized medicine, has yet been extended to account for cluster

randomization. This will impact p-values for tests of significance for individual differences but we do not expect impacts on the predictions themselves as parameter estimates are typically unbiased when clustering is ignored (Raudenbush & Bryk, 2002).

Methods

We use data collected as part of the cluster-randomized ACTION-PAC trial (Vallabhan et al., 2017), which included 8 high schools from a state in the Southwestern United States. Schools were eligible if they had functioning school-based health centers (SBHC), enrolled ≥700 students, had ≥40% Latinx students, and were located in high poverty areas. Each school was randomized to the intervention or control condition (see Figure 1). Participants in intervention schools were included in the intensive sample if their baseline BMI was ≥85th percentile, and in the prevention sample if their BMI was <85th percentile (Kuczmarski, 2002). Participants in control schools did not receive any intervention. The study ran from 2014 to 2017 and included 991 students in total, 608 in the prevention sample and 383 in the intensive treatment sample.

The parents of all students (intervention and control in both the intensive and prevention samples) received letters mailed home at baseline, midpoint (one year later), and endpoint (2 years later) with the child's health results. Letters outlined anthropometric measurements, blood pressure (BP) and cardiometabolic labs, highlighted normal or expected parameters for each marker, and healthy behaviors recommended by the American Academy of Pediatrics.

Students in the intensive sample at intervention schools received sixteen 20-minute sessions with a SBHC provider trained in motivational interviewing (MI) over two academic years. Students in the prevention sample at intervention schools received two 20-minute sessions from a SBHC provider over the same time frame. ACTION-PAC aimed to evaluate the effects of the intervention on reduction in BMI z-score for those in the intensive treatment sample, and

incidence of overweight and obesity for students in the prevention sample. Percentiles were calculated based on sex-specific growth charts for children ages 2-20 (Kuczmarski, 2002). Although PITE analyses are established for the binary outcomes targeted by the preventive intervention, only 8% of students in this sample were overweight by study endpoint, so power for the analyses would be greatly reduced by considering only the dichotomous threshold.

Participants

Participants were in the 9th or 10th grade. Consent was obtained from a parent and assent from the participant. Participants were excluded if they reported: diagnosis of type 1 or 2 diabetes; use of corticosteroids, antipsychotics, and/or medications to treat diabetes; inability to perform moderate to vigorous physical activity or were not ambulatory; hypertension, and/or hyperlipidemia; pregnancy; or developmental disorder(s) affecting weight or ability to understand study procedures. Participants were secondarily excluded if baseline assessments determined that their BP qualified as stage 2 hypertension or had a score ≥20 on the Eating Attitudes Test (EAT26), a screener to assess eating disorders (Garner et al., 1982).

In the prevention sample, 54.6% of participants were female and 85.4% were Latinx. The average age was 15.3 years (range: 13.4 years to 17.7 years), 95.6% of participants had a normal BMI and 4.4% were underweight (<5th percentile). In the intensive sample 55.1% of participants were female and 87.7% were Latinx. The average age was 15.3 years (range 14 to 17 years). At baseline, 51.2% had a BMI considered overweight (85th – 95th perentile), and 48.3% had a BMI considered obese (>95th percentile). Further, 10% were pre-diabetic (hemoglobin A1c [HgbA1c]: 5.7-6.4%), 27% had high triglycerides (≥130mg/dL), and 30% had low HDL cholesterol (<40mg/dL) (de Jesus, 2011).

Measures

Outcome

BMI percentiles for each participant were calculated based on height and weight at two-year follow-up. Percentiles were converted into BMI z-scores which was the primary outcome. BMI z-scores were used as the primary outcome in this analysis to align this study's aims with the original clinical trial study protocols. Previous research with adolescents has also used BMI-z to account for age and sex differences among adolescents (Wilson et al., 2015).

Baseline Predictors of Response to Treatment

Individual predictions are a function of the baseline covariates included in the predictive algorithm. In this study, the process for choosing covariates was conducted before any data were analyzed. One of the co-authors (EYJ) reviewed the available covariates in light of the biopsychosocial model (Suls et al., 2010) and chose those for which there was either theory or previous research suggesting that they might predict who would respond to treatment. The co-author was an investigator on the ActionPAC study, had previous knowledge of moderators of treatment effects, is experienced in interventions for prevention and treatment of obesity, and is clinically trained as a registered dietitian nutritionist.

All of the baseline variables described in Table 1 are included as predictors of individual differences in treatment response (treatment modifiers). The primary baseline variables used here included 45 variables for those in the prevention sample and 53 variables for those in the intensive treatment sample. The number differed between arms because some biomarkers were only assessed for participants in the intensive sample, per standard of care. However, for comparison we also conducted a secondary analysis for the intensive sample using only the 45 covariates which are also available in the prevention sample. Covariates in these analyses included individual-level demographics, biomarkers, dietary intake, accelerometry output, and

psycho-social measures. While PITEs can become inefficient and the permutation test loses power when too many covariates are included (Chang et al., 2021), the point of this method is to obtain predictions which take into account a broad array of factors that contribute to individual differences. Therefore, we suggest using theory and previous research to guide the selection of covariates which are likely to contribute to individual differences. Including multiple indicators of an imperfectly assessed construct, such as SES or dietary intake in the current sample, should improve predictions although at the cost of making it more difficult to detect if the variables make an important contribution to the predictions as the overall contribution will be shared across all related variables.

Demographic Covariates: Demographic covariates included in PITE analyses were sex, age, race/ethnicity, eligibility for free or reduced-price lunch, parents' level of education, occupation, and annual household income. Indicators of socioeconomic status were measured by parental report. Baseline covariates also included health history questions regarding family history of diabetes, heart attacks, and participation in weight management programs.

Biomarkers: Baseline values for BMI z-score, systolic and diastolic BP and waist circumference percentile, assessed by trained ACTION-PAC researchers, were included for both samples. Baseline fasting plasma glucose, hemoglobin A1C, insulin, total cholesterol, HDL and LDL cholesterol, triglycerides and HOMA-IR were also included for those in the intensive sample only as these biomarkers are typically assessed for students who are overweight.

Dietary Intake: All dietary intake variables were based on self-report using the 2007 version of the Block Food Screener for Ages 2-17, which has been validated in adolescents (Hunsberger et al., 2015). We included baseline measures of average daily intake of fruits and vegetables (measured in cups), added sugar (in teaspoons), as well as average daily glycemic

index and average daily glycemic load.

Physical Activity: Average baseline minutes of sedentary, moderate to vigorous, and vigorous activity as measured by accelerometer and the 3-Day Physical Activity Recall (3DPAR), which has been shown to be valid and reliable for adolescents, were included in the PITE analysis (Argiropoulou et al., 2004; Pate et al., 2003; Trost, 2007). Methods for collecting and processing physical activity measures are described in detail elsewhere (Sanders et al., 2019).

Psycho-Social Assessments: We included both some individual items and all subscales from the EAT-26 (Garner et al., 1982), a screening measure with established validity and reliability for identifying dieting behaviors, potential bulimia, and food preoccupation and oral control issues that may require further professional assessment (Garfinkel & Newman, 2001), and the Child Self-Report version (ages 13-18) of the 23-item Pediatric Quality of Life Inventory (PedsQL™), which has established reliability and validity as a measure of physical, emotional, social and school functioning in children and teens (Varni et al., 2001). The rationale for including items as well as the full scale is that there were some specific items which were expected to relate to treatment efficacy beyond the average of all items on the scale. Items included from the EAT-26 were assessed by frequency (0=”Never” to 5=”Often”) with which participants had (1) gone on eating binges where they felt like they might not be able to stop, (2) made themselves sick (vomited) to control their weight, and (3) used laxatives, diet pills, or diuretics to control their weight. Further, we included two individual items about motivation to change behavior. We included responses to the questions (1) “During the past 3 months, how ready have you felt to change your eating behavior?” and (2) “During the past 3 months, how ready have you felt to change your physical activity?” (0=“not at all ready” to 10=“extremely

ready”).

Data Analysis

To increase replicability of PITEs we believe that data analytic process should: 1) follow the original study protocol to the extent possible; 2) include a formal process for using theory and previous results for selecting covariates and predictive method; 3) rely on decisions made a priori to conducting PITE analyses; 4) include one test for heterogeneity in treatment effects; and 5) conduct both internal and external validation of model results. We also suggest that a protocol for the analyses proposed be written and registered in advance of testing for individual differences in a trial. We’ve described above the approach we took for selecting the outcomes and covariates. BART was selected as the predictive method prior to any analysis, based on the expectation of higher-order interactions which this method is well suited to detect.

To estimate PITEs in this analysis, we used BART (Chipman et al., 2010) as the primary predictive algorithm in steps 1 and 2, estimated with the BART package in R (McCulloch et al., 2019) using default settings for the priors. BART is a machine learning method which builds a series of regression trees with the tuning parameters chosen using Bayesian priors (Hill, 2011). A strength of BART is that it can detect higher-order interactions and non-linear effects, which are difficult to detect using a parametric model. In this case we used BART specifically because we expected higher order interactions between the baseline covariates in predicting treatment effects, whereas a linear model would only capture two-way interactions between the covariates and treatment. To check our coding and implementation of BART, we also computed PITEs using a linear regression model. In this case, PITEs from the two methods had a Pearson’s correlation of .65, suggesting that under half of the variability in the BART estimates could be attributed to two-way linear interactions between treatment and the covariates. The rest of the

variability is due to higher-order interactions and non-linear effects. Thus, in this case, much of the variation in the PITEs goes beyond the two-way interactions most likely to be included in moderation analyses.

After PITEs are estimated, but before the results are interpreted, it is important to show that the heterogeneity in treatment effects found is greater than what would be expected due to chance. Individual differences in treatment effects are quantified as the standard deviation (SD) of the PITEs across all individuals in the sample. If the treatment effect is predicted to be the same for everyone, the SD of the PITEs would be 0 (the null hypothesis). To test for the significance of individual differences we use a permutation test with 1000 bootstrap samples (Chang et al., 2021). This test has been shown to have adequate type I error rates and power given moderate effect sizes in previous simulation work (Chang et al., 2021).

In order to better understand which variables contribute to the individual differences observed in the effects of ACTION-PAC, our final analyses assessed variable importance (Bagherzadeh-Khiabani et al., 2016; Strobl et al., 2008) by re-estimating the PITEs separately for each covariate, removing that covariate from the full model and keeping all other variables. We examined the change in the SD of the PITEs as each variable was dropped from the full model. Similar to a backwards selection procedure, we consider the variable whose removal causes the largest decrease in the SD to have the largest impact on the PITE. We then removed that variable and restarted the iterative process, identifying the five variables responsible for the largest decreases in the PITE SD. These are the top candidates for explaining the individual differences in treatment effects observed. When estimated with BART, the impact of each variable is a combination of two-way interactions with treatment, and higher-order interactions, and non-linear interactions (Chipman et al., 2010). While the variable importance measure provides some

insight into which variables may be responsible for the individual difference in treatment effects, it is not intended or powered to show underlying mechanisms. A limitation of variable importance is that if multiple correlated predictors are included, each individual predictor is expected to be less likely to be identified as important than if one reliable predictor was included.

In this case, other decisions to be made a priori about data analysis include: 1) treatment of missing data; and 2) addressing clustering due to schools. Our a priori decision for missing data was to use single imputation with the "mice" package in R (van Buuren & Groothuis-Oudshoorn, 2011) to account for missingness, primarily in the baseline covariates. Single imputation was used because there is not yet an approach to adjust p-values for the permutation test using multiple imputation, thus the p-values presented for this test are somewhat liberal. We imputed data for the intensive and prevention samples separately. The imputation model included the same baseline covariates described above, as well as BMI z-score from both mid- and end-point. Although 3% of total scores were missing across all of the covariates and outcome data, 57% of the intensive and 48% of the prevention sample had at least one missing data point. After imputation, we used a regression predicting BMI which identified two outliers in the intensive sample with high influence on the predictions and for whom imputed BMI z-scores were higher than 5. Because influential datapoints change predictions for everyone, we made the decision before conducting PITE analyses to exclude these cases. In retrospect, a better decision would have been to use an imputation model which reduced the likelihood of influential datapoints (such as predictive mean matching), we report the original results here in order to avoid making decisions after the original results were obtained which could impact the replicability of those results.

In this study, clustering due to school is very important as randomization occurred at the

level of school. We made the decision to ignore clustering here, since this is intended as a demonstration of PITE and because PITE has yet to be extended to incorporate clustered data. Addressing clustering inside of predictive models is especially difficult because the predictive algorithm needs to apply to new cases from different clusters and we don't know what the cluster level effects are for those. The impact of ignoring clustering is expected to be primarily on the p-values for the permutation test because estimates obtained from models which ignore cluster-level randomization are generally unbiased (Raudenbush & Bryk, 2002). If randomization had been at the individual level we see no reason that clustering would typically need to be accounted for.

Results

We start by examining the distributions of the ACTION-PAC baseline covariates and outcomes for both the prevention and intensive treatment samples. Table 1 shows descriptive statistics for each sample.

Individual differences in the effects of intervention among the intensive sample

The intensive sample of ACTION-PAC participants included 381 individuals, 184 in treatment schools and 197 in control schools. The SD of the PITEs for this sample was 0.35. The permutation test showed that the expected SD of the PITEs, given no heterogeneity in treatment effects, is 0.32. The p-value is above .05, thus, individuals in the intensive sample did not respond significantly differently to the intervention as a function of the covariates we examined. Given the results of the permutation test, we do not further describe the PITE estimates for students in the intensive treatment condition. A secondary analysis to allow the results from the intensive and preventive interventions to be compared was run in which we computed PITEs and ran the permutation test using only the same 45 baseline predictors available for both samples,

the permutation test was still not significant in the intensive group and the standard deviation of the PITEs was .279.

Individual differences in treatment effects in the prevention sample

The prevention sample of ACTION-PAC participants consisted of 608 individuals, 318 in treatment schools and 290 in control schools. The SD of the PITE for students in the prevention sample was 0.39. The probability of finding this result given no individual differences in the effect of the intervention was 0.04, suggesting significant heterogeneity in the effects of treatment given the baseline covariates.

For the prevention sample, PITEs ranged from -1.36 to 1.39, with a median predicted value of -0.13. The median is a non-parametric estimate of the total treatment effect on BMI z-score, conditioning on all of the covariates in the analysis. This indicates there was a small difference in BMI z-score in the treatment group versus the control group. Average treatment effects were not tested here and will be reported in separate papers using methods proposed in the initial study grant. The distribution of the PITEs (see Figure 2) shows that 65.3% of the sample had a PITE score (BMI-z scale) less than zero. In this context, a predicted value of less than zero indicates that an individual's baseline characteristics predict they would experience a reduction in BMI z-score in response to the ACTION-PAC intervention. The PITE distribution shows that 25% of students were expected to have BMI z-scores reduced by 0.39 or greater by the preventive intervention. Conversely, 25% of those in the prevention sample are predicted to have their BMI z-score increased by 0.09 or more. Note that this does not account for uncertainty in the predictions. For any particular individual, the difference between treatment and control might not be significant.

Variables which contribute to individual differences in the preventive intervention

Neither treatment nor control groups showed any average changes in BMI z-score across the course of the study (change in BMI-z was ± .01 in both groups). We first examined the correlation between BMI z-score at baseline and predicted treatment effects. These were non-significantly correlated at 0.14. Baseline BMI z-score was not related to predicted treatment effects. To better understand the variables contributing to observed differences in the effect of the intervention we conducted a test of variable importance by removing each baseline covariate, one at a time, and assessing model changes. We note that the decreases in SD that we observed were small and it is likely that the variables identified could be different in a different sample. We also note that, unlike the predictions, this test is likely to be quite sensitive to the inclusion of correlated predictors.

Variable importance results (see Table 2) show that amount of added sugar in a participant's diet contributed the most to individual differences in responses to the intervention. This was followed by social quality-of-life (QOL) scale score, emotional QOL scale score, readiness to change physical activity, and having had eating binges and feeling unable to stop. To understand the relationship between these variables and PITEs, we ran correlations. Higher amounts of added sugar in one's diet was associated with a lower PITE score, indicating a higher predicted effect on BMI z-score. Higher social and emotional QOL scale scores were also associated with lower PITE scores. In other words, individuals with high added sugar in their diet and low levels of social and emotional difficulty at baseline benefitted the most from the intervention.

On the other hand, higher baseline values for motivation to change physical activity and frequency of binge eating were associated with higher PITE scores. This positive correlation indicates readiness to change physical activity and frequent binging are associated with lower

effectiveness of the intervention. Because motivational interviewing comprised a key part of intervention, it is logical that those low in motivation at baseline would benefit more from receiving the intervention than those who already had high motivation to change.

Discussion

While strong arguments have been made for using personalized medicine to target interventions for the treatment or prevention of obesity, few studies have tested the use of new personalized medicine methods in the area. Our analysis uses statistical approaches and a machine learning algorithm to demonstrate that a data-driven approach to personalized medicine, based on the biopsychosocial model (Suls et al., 2010), can identify individual differences in the effects of a psychosocial intervention targeting adiposity. To our knowledge this is the first study to demonstrate the use of these methods for personalized medicine to identify individual differences in the effects of treatments targeting adiposity.

We only showed significant individual differences for those receiving the prevention intervention and these results are qualified because we could not account for school-level randomization or perform multiple imputation. If replicated, these differences would be notable because of the very low intensity of the intervention (only 20 min per academic year per student). Although there was no significant average treatment effect, these results suggest that this very low intensity treatment could help some students more than others. That said, the primary objective of personalized medicine methods is to obtain predictions which can improve outcomes by helping clinicians choose the best treatment. The potential impact on clinical care would have been much clearer if we were able to predict response to the intensive intervention for clients already overweight or obese. In terms of effects for the intensive sample, recent research suggests that power to find effects for this sample was likely limited (Chang et al.,

2021). A larger sample would have been needed to conclude that there were not individual differences in this intervention.

These results demonstrate the possibility of using the PITE framework to find and describe individual heterogeneity in the effects of an obesity intervention. A major point of this paper is that, from the perspective of personalized medicine, what ultimately matters is not the specific mechanisms but whether clinically useful predictions can be made. Predictive models are best constructed with careful attention to theory and previous results (Hoogland et al., 2021). However, the strength of personalized medicine methods is their potential for clinical application rather than the ability to test theory (Kuhlemeier et al., 2021). While the clinical utility of this example would have increased if effects had been found for the intensive intervention, Figure 2 shows how the method results in a prediction about the treatment effect for each individual. Ultimately, translating these approaches into clinical practice would be facilitated by web based collection of client data, automated calculation of the client's prediction and predictive interval, and providing this data to the clinician and their client in a user—friendly format. External evidence for the validity of these predictions would be needed to justify clinical applications.

We see the primary contribution of this paper as demonstrating that personalized medicine methods have potential to identify individuals for whom the effect of a behavioral treatment is likely to be larger or smaller than the average effect. One of the things this paper highlights is that personalized medicine methods need to be extended in multiple ways to account for situations seen in real world clinical trials. First, at present the methodology is unable to account for clustering which is needed for group randomized trials. Methods have been proposed for permutation tests with clustered data (Braun & Feng, 2001) so the overall test of individual differences should be achievable. More problematic is obtaining predictions for

individuals from new clusters (schools in our example). Second, methods for estimating individual-level predictive intervals with non-parametric predictive methods need to be established. Predictive intervals make it easier to interpret individual results and use PITEs for choosing between treatments. Third, methods for using multiply imputed data and for obtaining new predictions in the presence of missing data are badly needed. Fourth, new work on validation methods for PITEs is needed. Methods for validation of results would provide a tool for verifying that predictions from one sample are likely to generalize and would provide a tool for understanding what went wrong if these predictions are not replicated. A reviewer suggested presenting the accuracy of the predictions under each condition. While intuitive, this approach is problematic because predictions under each condition capture main effects as well as differences in the effect of treatment. As PITEs are never observed, measures of predictive accuracy of the PITE are not available. Finally, the goal of using PITEs is to choose the intervention most likely to be effective for a particular patient. Achieving this goal requires predictions which provide comparisons of multiple treatments. Future personalized medicine research should prioritize such comparisons.

One of the contributions of this study is the lessons to be learned for future use of personalized medicine methods. As was made clear by our choice of imputation methods and the use of BMI-z as an outcome, initial decisions made in the process can be consequential and open to questioning. Nevertheless, the ultimate goal of personalized medicine is to obtain predictions which will replicate in new data, a goal which is facilitated by making analytic decisions a priori. We propose several steps: 1) care should be taken to anticipate and plan for issues likely to arise in analyses, such as missing data and outliers, a priori and screening of baseline data should be conducted prior to estimating predictive algorithms; 2) the design of the original clinical trial

should be mirrored as much as possible (for example, in using BMI-z as the outcome); 3) approaches for increasing replicability such as pre-publication of protocols of studies implementing personalized medicine approaches are recommended; and 4) external validation of predictions in new samples provides the strongest evidence to support ultimate implementation in clinical settings. Another lesson learned is that while in the current study we had an investigator select covariates a priori based on theory and knowledge of the field, this process could be further improved by having a panel of experts arrive at a consensus informed by formal literature review. This panel should be instructed on how to balance being inclusive in choosing covariates so as not to miss individual differences, with including so many covariates that sampling noise is increased and power is reduced. Another alternative is new methods for reducing dimensionality of the covariates while maintaining higher order interactions. We also note that using PITE predictions for targeting a low-intensity prevention intervention may not be practical, as it would require obtaining data on students in order to make predictions. These methods are more useful when applied to clinical populations.

Advancement in personalized medicine methods would also be facilitated by consideration of treatment compliance as an additional outcome. Application of the PITE method could be applied to determine whether certain baseline characteristics predict meaningful heterogeneity in compliance with a given intervention. The results of studies such as these could enable interventionists to determine whether a particular patient, on the basis of a group of baseline characteristics, would be likely to comply with a given treatment regimen.

In sum, this study was intended primarily to demonstrate the potential of methods for personalized medicine with interventions targeting adiposity. To date, many have discussed the potential of personalized medicine in this area, but few papers are published showing that it is

possible to predict individual's treatment response. This study both shows the promise that this approach can have while also illustrating that much work needs to be done in the area. This study also illustrates that the strength of these methods is in their predictions rather than their ability to inform theory about underlying mechanisms.

References

Angrist, J. D., Imbens, G. W., Rubin, D. B., Angrist, J. D., Imbens, G. W., Identification, D. B. R., Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, *91*(434), 444–455.

Argiropoulou, E. C., Michalopoulou, M., Aggeloussis, N., & Avgerinos, A. (2004). Validity and reliability of physical activity measures in greek high school age children. *Journal of Sports Science and Medicine*, *3*(3), 147–159.

Bagherzadeh-Khiabani, F., Ramezankhani, A., Azizi, F., Hadaegh, F., Steyerberg, E. W., & Khalili, D. (2016). A tutorial on variable selection for clinical prediction models: feature selection methods in data mining could improve the results. *Journal of Clinical Epidemiology*, *71*, 76–85. https://doi.org/10.1016/j.jclinepi.2015.10.002

Ballarini, N. M., Rosenkranz, G. K., Jaki, T., Konig, F., & Posch, M. (2018). Subgroup identification in clinical trials via the predicted individual treatment effect. *PLOS One*, *13*(10), 1–22. https://doi.org/10.1371/journal.pone.0205971

Baranowski, T., Motil, K. J., & Moreno, J. P. (2019). Multi-etiological Perspective on Child Obesity Prevention. *Current Nutrition Reports*, *8*(1), 1–10. https://doi.org/10.1007/s13668-019-0256-3

Beets, M. W., Brazendale, K., Weaver, R. G., & Armstrong, B. (2019). Rethinking Behavioral Approaches to Complement Biological Advances to Understand the Etiology, Prevention, and Treatment of Childhood Obesity. *Childhood Obesity*, *15*(6), 353.

Braun, T. M., & Feng, Z. (2001). Optimal permutation tests for the analysis of group randomized trials. *Journal of the American Statistical Association*, *96*(456), 1424–1432.

https://doi.org/10.1198/016214501753382336

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324

Chang, C., Jaki, T., Sadiq, M. S., Kuhlemeier, A., Feaster, D. J., Cole, N., Lamont, A., Oberski, D., Desai, Y., & Van Horn, M. L. (2021). A permutation test for assessing the presence of individual differences in treatment effects. *Statistical Methods in Medical Research, Avl Online*(September). https://doi.org/10.1177/09622802211033640

Chatelan, A., Bochud, M., & Frohlich, K. L. (2019). Precision nutrition: Hype or hope for public health interventions to reduce obesity? *International Journal of Epidemiology*, *48*(2), 332–342. https://doi.org/10.1093/ije/dyy274

Chipman, H. A., George, E. I., & McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *Annals of Applied Statistics*, *6*(1), 266–298. https://doi.org/10.1214/09-AOAS285

de Jesus, J. M. (2011). Expert Panel on Integrated Guidelines for Cardiovascular Health and Risk Reduction in Children and Adolescents: Summary Report. *Pediatrics*, *128*, S213–S256. https://doi.org/10.1542/peds.2009-2107C

Estampador, A. C., & Franks, P. W. (2018). Precision Medicine in Obesity and Type 2 Diabetes : The Relevance of Early-Life Exposures. *Clinical Chemistry*, *64*(1), 130–141. https://doi.org/10.1373/clinchem.2017.273540

Frühbeck, G., Kiortsis, D. N., & Catalan, V. (2018). Precision medicine : diagnosis and management of obesity. *The Lancet*, *6*, 164–166. https://doi.org/10.1016/S2213-8587(17)30312-1

Garfinkel, P. E., & Newman, A. (2001). *The Eating Attitudes Test: Twenty-five years later*.

Garner, D. M., Olmsted, M. P., Bohr, Y., & Garfinkel, P. E. (1982). The Eating Attitudes Test:

Psychometric features and clinical correlates. *Psychological Medicine*, *12*, 871–878.

Green, D. P., & Kern, H. L. (2012). Modeling heterogeneous treatment effects in survey experiments with bayesian additive regression trees. *Public Opinion Quarterly*, *76*(3), 491–511. https://doi.org/10.1093/poq/nfs036

Hales, C. M., Fryar, C. D., Carroll, M. D., Freedman, D. S., & Ogden, C. L. (2018). Trends in Obesity and Severe Obesity Prevalence in US Youth and Adults by Sex and Age , 2007-2008 to 2015-2016. *JAMA - Journal of the American Medical Association*, *319*(16), 1723–1725.

Henderson, N. C., Louis, T. A., & Rosner, G. L. (2017). *Individualized Treatment Effects with Censored Data via Fully Nonparametric Bayesian Accelerated Failure Time Models*. 1–41.

Heymsfield, S., Aronne, L. J., Eneli, I., Kumar, R. B., Michalsky, M., Walker, E., Wolfe, B. M., Woolford, S. J., & Yanovski, S. (2018). Clinical Perspectives on Obesity Treatment: Challenges, Gaps, and Promising Opportunities. *NAM Perspectives*, *Discussion*. https://doi.org/10.31478/201809b

Hill, J. L. (2011). Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, *20*(1), 217–240. https://doi.org/10.1198/jcgs.2010.08162

Hoelscher, D. M., Kirk, S., Ritchie, L., & Cunningham-Sabo, L. (2013). Position of the Academy of Nutrition and Dietetics: Interventions for the Prevention and Treatment of Pediatric Overweight and Obesity. *Journal of the Academy of Nutrition and Dietetics*, *113*(10), 1375–1394. https://doi.org/10.1016/j.jand.2013.08.004

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*(396), 945–960.

Hoogland, J., IntHout, J., Belias, M., Rovers, M. M., Riley, R. D., Harrell, F. E., Moons, K. G. M., Debray, T. P. A., & Reitsma, J. B. (2021). A tutorial on individualized treatment effect prediction from randomized trials with binary endpoint. *Statistics in Medicine, Online*.

Hunsberger, M., O'Malley, J., Block, T., & Norris, J. C. (2015). Relative validation of Block Kids Food Screener for dietary assessment in children and adolescents. *Maternal and Child Nutrition, 11*, 260–270. https://doi.org/10.1111/j.1740-8709.2012.00446.x

Imai, K., & Strauss, A. (2011). Estimation of heterogeneous treatment effects from randomized experiments, with application to the optimal planning of the get-out-the-vote campaign. *Political Analysis, 19*(1), 1–19. https://doi.org/10.1093/pan/mpq035

Ingelsson, E., & McCarthy, M. I. (2018). Human Genetics of Obesity and Type 2 Diabetes Mellitus. *Circulation: Genomic and Precision Medicine, 11*, 1–12. https://doi.org/10.1161/CIRCGEN.118.002090

Irvin, V. L., & Kaplan, R. M. (2016). Effect Sizes and Primary Outcomes in Large-Budget, Cardiovascular-Related Behavioral Randomized Controlled Trials Funded by NIH Since 1980. *Annals of Behavioral Medicine, 50*, 130–146. https://doi.org/10.1007/s12160-015-9739-7

Kelly, A. S., Marcus, M. D., Yanovski, J. A., Yanovski, S. Z., & Osganian, S. K. (2018). Working toward precision medicine approaches to treat severe obesity in adolescents: report of an NIH workshop. *International Journal of Obesity*, *42*, 1834–1844. https://doi.org/10.1038/s41366-018-0231-x

Kent, D. M., Steyerberg, E., & van Klaveren, D. (2018). Personalized evidence based medicine : predictive approaches to heterogeneous treatment effects. *BMJ, 364*, k4245. https://doi.org/10.1136/bmj.k4245

Kranzler, H. R., & McKay, J. R. (2012). Personalized Treatment of Alcohol Dependence. *Substance Use and Related Disorders*, *14*, 486–493. https://doi.org/10.1007/s11920-012-0296-5

Kuczmarski, R. J. (2002). *2000 CDC Growth Charts for the United States: methods and development* (Issue 246). Department of Health and Human Services, Centers for Disease Control and ….

Kuhlemeier, A., Desai, Y., Tonigan, A. A., Witkiewitz, K., Jaki, T., Hsiao, Y. Y., Chang, C., & Van Horn, M. L. (2021). Applying Methods for Personalized Medicine to the Treatment of Alcohol Use Disorder. *Journal of Consulting and Clinical Psychology*, *89*(4), 288.

Lamont, A., Lyons, M. D., Jaki, T., Stuart, E., Feaster, D. J., Tharmaratnam, K., Oberski, D., Ishwaran, H., Wilson, D. K., & Van Horn, M. L. (2018). Identification of predicted individual treatment effects in randomized clinical trials. *Statistical Methods in Medical Research*, *27*(1), 142–157. https://doi.org/10.1177/0962280215623981

Lemstra, M., Bird, Y., Nwankwo, C., Rogers, M., & Moraros, J. (2016). Weight loss intervention adherence and factors promoting adherence: a meta-analysis. *Patient Preference and Adherence*, *10*, 1547–1559.

MacLean, P. S., Rothman, A. J., Nicastro, H. L., Czajkowski, S. M., Agurs-Collins, T., Rice, E. L., Courcoulas, A. P., Ryan, D. H., Bessesen, D. H., & Loria, C. M. (2018). The Accumulating Data to Optimally Predict Obesity Treatment (ADOPT) Core Measures Project : Rationale and Approach. *Obesity*, *26*, S6–S15. https://doi.org/10.1002/oby.22154

McCulloch, R., Sparapani, R., Gramacy, R., Spanbauer, C., & Pratola, M. (2019). *BART: Bayesian Additive Regression Trees.* R package version 2.2. https://cran.r-project.org/package=BART%0A

O'Connor, E. A., Evans, C. V, Burda, B. U., Walsh, E. S., Eder, M., & Lozano, P. (2017). Screening for Obesity and Intervention for Weight Management in Children and Adolescents: Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA - Journal of the American Medical Association*, *317*(23), 2427–2444. https://doi.org/10.1001/jama.2017.0332

Pate, R. R., Ross, R., Dowda, M., Trost, S. G., & Sirard, J. R. (2003). Validation of a 3-Day Physical Activity Recall Instrument in Female Youth. *Pediatric Exercise Science*, *15*, 257–265.

Poulson, R. S. (2011). Treatment heterogeneity and individual qualitative interaction. In *ProQuest Dissertations and Theses*. Kansas State University.

Preston, S. H., Vierboom, Y. C., & Stokes, A. (2018). The role of obesity in exceptionally slow US mortality improvement. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(5), 957–961. https://doi.org/10.1073/pnas.1716802115

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (Second). Sage Publications.

Rubin, D. B. (2005). Causal Inference Using Potential Outcomes Causal Inference Using Potential Outcomes : Design , Modeling , Decisions. *Journal of the American Statistical Association*, *100*(469), 322–331. https://doi.org/10.1198/016214504000001880

Ryder, J. R., Kaizer, A. M., Jenkins, T. M., Kelly, A. S., Inge, T. H., & Shaibi, G. Q. (2019). Heterogeneity in Response to Treatment of Adolescents with Severe Obesity: The Need for Precision Obesity Medicine. *Obesity*, *27*(2), 288–294. https://doi.org/10.1002/oby.22369

Severin, R., Sabbahi, A., Mahmoud, A. M., Arena, R., & Phillips, S. A. (2019). Progress in Cardiovascular Precision Medicine in Weight Loss and Healthy Living. *Progress in*

Cardiovascular Diseases, *62*(1), 15–20. https://doi.org/10.1016/j.pcad.2018.12.012

Smith, M. D., Saunders, R. S., Stuckhardt, L., & McGinnis, J. M. (2013). *Best care at lower cost: The path to continuously learning health care in America* (Institute of Medicine (ed.)). National Academies Press.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, *9*, 307–318. https://doi.org/10.1186/1471-2105-9-307

Suls, J. M., Luger, T., & Martin, R. (2010). The Biopsychosocial Model and the Use of Theory in Health Psychology. In J. M. Suls, K. W. Davidson, & R. M. Kaplan (Eds.), *Handbook of Health Psychology and Behavioral Medicine*. The Guilford Press.

Trost, S. G. (2007). State of the Art Reviews: Measurement of Physical Activity in Children and Adolescents. *American Journal of Lifestyle Medicine*, *1*(4), 299–314. https://doi.org/10.1177/1559827607301686

Vallabhan, M. K., Kong, A. S., Yakes-Jimenez, E., Summers, L. C., DeBlieck, C. J., & Feldstein Ewing, S. W. (2017). Training Primary Care Providers in the Use of Motivational Interviewing for Youth Behavior Change. *Research and Theory for Nursing Practice*, *31*(3), 219–232. https://doi.org/10.1891/1541-6577.31.3.219.Training

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). *mice: Multivariate Statistical, Imputation by Chained Equations in R*. Journal of Software, 45(3), 1-67. https://www.jstatsoft.org/v45/i03/.

Varni, J. W., Seid, M., & Kurtin, P. S. (2001). PedsQL™ 4.0: Reliability and Validity of the Pediatric Quality of Life Inventory™ Version 4.0 Generic Core Scales in Healthy and Patient Populations. *Medical Care*, *39*(8), 800–812. https://doi.org/10.1097/00005650-

200108000-00006

Wang, Y., Cai, L., Wu, Y., Wilson, R., Weston, C., Fawole, O., Bleich, S., Cheskin, L., Showell, N., Lau, B., Chiu, D., Zhang, A., & Segal, J. (2015). What childhood obesity prevention programmes work? A systematic review and meta-analysis. *Obesity Reviews*, *16*(7), 547–565. https://doi.org/10.1111/obr.12277.What

Webb, C. A., Cohen, Z. D., Beard, C., Forgeard, M., Peckham, A. D., & Björgvinsson, T. (2020). Personalized Prognostic Prediction of Treatment Outcome for Depressed Patients in a Naturalistic Psychiatric Hospital Setting : A Comparison of Machine Learning Approaches. *Journal of Consulting and Clinical Psychology*, *88*(1), 25–38.

Wilson, D. K., Kitzman-Ulrich, H., Resnicow, K., Van Horn, M. L., George, S. M. S., Siceloff, E. R., Alia, K. A., McDaniel, T., Heatley, V. S., Huffman, L., Coulon, S., & Prinz, R. (2015). An overview of the Families Improving Together (FIT) for weight loss randomized controlled trial in African American families. *Contemporary Clinical Trials*, *42*, 145–157. https://doi.org/10.1016/j.cct.2015.03.009

Xiaotong, J., Nelson, A. E., Cleaveland, R. J., Beavers, D. P., Schwartz, T. A., Arbeeva, L., Alvarez, C., Callahan, L. F., Messier, S., Loeser, R. F., & Kosorok, M. R. (2020). *A Precision Medicine Approach to Develop and Internally Validate Weight Loss Treatments for Overwieght and Obese Adults with Knee Oesteoarthritis*. 0–3. https://doi.org/10.1002/acr.24179

Yanovski, S. Z., & Yanovski, J. A. (2018). Toward Precision Approaches for the Prevention and Treatment of Obesity. *JAMA - Journal of the American Medical Association*, *319*(3), 223–224.

Table 1

Descriptive Statistics by Sample for Covariates Used to Develop PITEs

	Intensive Sample						Prevention Sample					
	Intervention n=184			Control n=197			Intervention n=318			Control n=290		
Variables in PITE analyses	Mean	IQR	Miss	Mean	IQR	Miss	Mean	IQR	Miss	Mean	IQR	Miss
Endpoint BMI z-score	1.72	[1.3, 2.1]	65	1.70	[1.2, 2.2]	67	0.06	[-0.5, 0.8]	90	-0.16	[-0.8, 0.6]	83
Demographic												
Male	46%		0	45%		0	46 %		0	45%		0
Age (years)	15.3	[14.7, 15.8]	0	15.4	[14.8, 15.9]	0	15.3	[14.8, 15.8]	0	15.4	[14.8,15.9]	0
Parental Education												
Less than high school	29%		3	32%		1	35%		1	27%		3
High school graduate	27%		3	27%		1	21%		1	22%		3
Some college	30%		3	30%		1	31%		1	27%		3
College graduate	14%		3	11%		1	13%		1	22%		3
Parental Work Status												
Full-time	51%		2	44%		1	51%		4	38%		3
Part-time	9%		2	10%		1	10%		4	14%		3
Self-Employed	7%		2	8%		1	7%		4	10%		3
Out of Work	17%		2	22%		1	19%		4	26%		3
No Need/Unknown	16%		2	16%		1	13%		4	10%		3
Race/Ethnicity												
Latinx	92%		0	83%		0	88%		0	83%		0
White	10%		0	13%		0	10%		0	16%		0
Black	2%		0	5%		0	4%		0	4%		0
American Indian	2%		0	4%		0	3%		0	2%		0
Eligible for free and reduced lunch	86%		6	82%		5	82%		6	76%		0
Household income	3.39	[2.0, 5.0]	7	3.28	[2.0, 4.2]	5	3.45	[2.0, 5.0]	10	3.53	[2.0, 5.0]	13
Scales												
EAT-26 Items & Subscales												
Frequency of eating binges	1.71	[1.0, 2.0]	0	1.61	[1.0, 2.0]	0	1.59	[1.0, 2.0]	0	1.54	[1.0, 2.0]	0
Frequency of vomiting	1.09	[1.0, 1.0]	0	1.10	[1.0, 1.0]	0	1.05	[1.0, 1.0]	0	1.02	[1.0, 1.0]	0
Frequency of laxative use	1.17	[1.0, 1.0]	0	1.10	[1.0, 1.0]	0	1.02	[1.0, 1.0]	0	1.02	[1.0, 1.0]	0
Dieting subscale	5.30	[2.0, 8.0]	0	5.27	[2.0, 8.0]	0	3.12	[1.0, 4.0]	0	2.78	[1.0, 3.8]	0
Bulimia/Food preoccupation subscale	0.38	[0.0, 0.0]	0	0.32	[0.0, 0.0]	0	0.36	[0.0, 0.0]	0	0.45	[0.0, 0.0]	0
Oral Control subscale	2.01	[0.0, 3.0]	0	1.82	[0.0, 3.0]	0	2.30	[1.0, 3.0]	0	2.40	[1.0, 3.0]	0
Readiness to change diet	6.14	[4.0, 8.0]	0	5.94	[5.0, 8.0]	0	4.40	[2.0, 6.0]	0	4.06	[1.0, 6.0]	0
Readiness to change PA	6.72	[5.0, 9.0]	0	6.77	[5.0, 9.0]	0	5.67	[4.0, 8.0]	0	5.61	[4.0, 8.0]	0

PedsQL™												
Physical QoL	83.5	[75.0, 93.8]	0	84.8	[78.1, 93.8]	0	84.5	[78.1, 93.8]	1	85.5	[81.3, 93.8]	0
Social QoL	84.5	[75.0, 95.0]	0	84.8	[75.0, 95.0]	0	87.3	[80.0, 100.0]	0	86.2	[80.0,100.0]	0
Emotional QoL	74.6	[60.0, 90.0]	0	72.6	[60.0, 90.0]	1	74.3	[60.0, 90.0]	1	75.9	[65.0,90.0]	1
School QoL	68.5	[55.0, 80.0]	1	66.4	[55.0, 75.0]	0	70.0	[60.0, 85.0]	0	71.2	[60.0,85.0]	0
Medical and Family History												
Baseline BMI z-score	1.68	[1.3, 2.0]	0	1.71	[1.3, 2.1]	0	0.05	[-0.3, 0.6]	0	-0.09	[-0.6, 0.5]	0
Systolic blood pressure z-score	0.05	[-0.5, 0.5]	0	0.17	[-0.3, 0.7]	0	-0.59	[-1.2, -0.1]	0	-0.60	[-1.2, -0.0]	0
Diastolic blood pressure z-score	-0.11	[-0.5, 0.3]	0	-0.01	[-0.5, 0.5]	0	-0.40	[-0.8, 0.0]	0	-0.30	[-0.8, 0.2]	0
Waist Circumference percentile	88.3	[83.7, 94.8]	0	88.4	[84.2, 95.4]	1	47.1	[31.1, 64.7]	2	44.5	[29.2, 61.1]	1
Sees doctor for medical condition	21%		5	28%		4	21%		9	16%		8
Takes medication	19%		4	24%		4	23%		8	19%		7
Family with Type II Diabetes	54%		15	60%		7	47%		0	42%		1
Immediate family with Type II diabetes	10%		0	14%		0	8%		0	4%		0
Family heart attack before age 55	23%		12	28%		15	15%		13	15%		20
Immediate family, heart attack	2%		0	3%		0	2%		0	1%		0
Weight loss program, last year	10%		0	5%		0	2%		1	2%		0
Dietary Variables												
Fruit (cups)	1.4	[0.6, 1.9]	1	1.4	[0.5, 1.9]	1	1.4	[0.7, 2.0]	1	1.5	[0.6, 2.1]	1
Vegetables, no potatoes (cups)	0.7	[0.3, 0.9]	1	0.7	[0.3, 0.8]	1	0.7	[0.4, 0.9]	1	0.8	[0.4, 1.0]	1
Added sugar (tsp)	6.7	[3.3, 8.1]	1	7.0	[3.1, 8.6]	1	9.3	[3.9, 11.6]	1	9.0	[4.6, 11.3]	1
Glycemic index	48.8	[46.1, 51.4]	1	48.6	[45.5, 51.0]	1	49.0	[46.3, 51.0]	1	49.1	[46.6, 51.5]	1
Glycemic load	61.9	[38.9, 75.7]	1	61.7	31.9	1	78.0	[47.3, 92.5]	1	76.5	[48.3, 97.2]	1
Physical Activity												
3-Day PAR												
Average energy expended	73.8	[64.7, 82.4]	28	74.58	[63.1, 84.0]	14	78.0	[64.0, 82.6]	46	75.18	[64.5, 89.6]	34
Average vigorous PA blocks	1.9	[0.3, 3.0]	28	1.9	[0.0, 3.3]	14	2.4	[0.0, 4.0]	46	2.0	[0.0, 3.0]	34
Average MVPA blocks	5.5	[3.3, 7.3]	28	5.9	[3.3, 8.3]	14	6.0	[3.3, 8.3]	46	5.4	[3.0, 7.3]	34
Average sedentary PA blocks	29.5	[27.0, 31.3]	28	29.3	[26.7, 31.7]	14	28.8	[26.3, 31.7]	46	29.5	[27.0, 32.3]	34
Accelerometer (minutes)												
Weekly avg sedentary PA	801.3	[749.3 847.8]	9	818.3	[757.4 871.6]	17	799.5	[748.6, 847.8]	27	811.4	[759.2, 870.4]	19
Weekly avg MVPA	53.3	[30.8, 69.5]	9	51.6	[28.6, 69.0]	17	56.3	[31.2, 77.1]	27	53.6	[29.5, 72.5]	19
Weekly avg vigorous PA	4.1	[0.6,	9	4.0	[0.5,	17	6.3	[1.0,	27	5.5	[0.8,	19

		5.6]			5.3]		8.2]	7.7]
Lab Data								
Glucose (mg/dL)	90.3	[85.0, 95.0]	5	89.9	[85.0, 95.0]	4		
HgbA1C (%)	5.2	[5.0, 5.5]	5	5.3	[5.2, 5.5]	4		
Insulin (uIU/mL)	21.6	[13.8, 26.0]	5	24.3	[13.0, 29.0]	4		
Cholesterol (mg/dL)	142.0	[123.0 154.3]	5	146.0	[126.0 164.0]	4		
HDL (mg/dL)	44.4	[38.0, 43.5]	5	45.9	[39.0, 52.0]	4		
LDL (mg/dL)	75.7	[58.0, 87.5]	5	77.90	[64.0, 93.0]	5		
Trigylcerides (mg/dL)	109.6	[66.0, 135.3]	5	110.9	[70.0, 137.0]	4		
HOMAIR (mg/dL)	4.9	[2.9, 5.9]	5	5.46	[2.9, 6.4]	4		

Table 2

Variable Importance by Largest Reduction in Standard Deviation of the PITE (SD of full model =.39)

		Standard Deviation of PITE			
Baseline Variables	Full Model	-1 variable	-2	-3	-4
Most Important Variable	sugar added	emotional QoL	social QoL	readiness to change	binge frequency
Male	0.380	0.379	0.361	0.348	0.342
EAT-26 Items & Subscales					
Frequency of eating binges	0.377	0.359	0.360	0.345	**0.330**
Frequency of vomiting	0.373	0.359	0.352	0.337	0.341
Frequency of laxative use	0.379	0.370	0.362	0.342	0.344
Dieting subscale	0.373	0.358	0.364	0.350	0.338
Bulimia/Food preoccupation subscale	0.382	0.358	0.361	0.348	0.346
Oral Control subscale	0.376	0.361	0.364	0.346	0.342
Readiness to change diet	0.370	0.364	0.360	0.345	0.339
Readiness to change physical activity	0.365	0.355	0.359	**0.335**	
PedsQL™					
Physical QoL	0.373	0.363	0.367	0.358	0.338
Social QoL	0.384	0.366	**0.352**		
Emotional QoL	0.366	**0.336**			
School QoL	0.373	0.365	0.371	0.354	0.341
Age	0.368	0.354	0.353	0.341	0.335
Baseline BMI z-score	0.386	0.391	0.388	0.381	0.365
Systolic blood pressure z-score	0.370	0.364	0.357	0.347	0.338
Diastolic blood pressure z-score	0.372	0.355	0.369	0.347	0.342
Waist Circumference percentile	0.383	0.361	0.361	0.352	0.341
Sees doctor for medical condition	0.382	0.362	0.369	0.360	0.353
Takes medication for medical condition	0.380	0.367	0.359	0.356	0.345
Anyone in family with Type II Diabetes	0.386	0.365	0.367	0.350	0.351
Immediate family with Type II Diabetes	0.372	0.375	0.365	0.360	0.342
Anyone in family, heart attack before age 55	0.369	0.360	0.357	0.350	0.330
Immediate family, heart attack	0.383	0.362	0.367	0.352	0.348
Involved in weight loss program, last 12 months	0.387	0.361	0.369	0.351	0.341
Eligible for free & reduced lunch	0.372	0.367	0.359	0.347	0.338
Household income	0.383	0.377	0.367	0.352	0.341
Parental education	0.374	0.362	0.370	0.348	0.337
Parental work status	0.376	0.362	0.366	0.352	0.345
Race/Ethnicity					

Hispanic	0.377	0.376	0.374	0.349	0.350
White	0.371	0.359	0.369	0.348	0.348
Black	0.379	0.364	0.367	0.348	0.347
American Indian	0.376	0.365	0.368	0.348	0.338
Fruit (cups)	0.379	0.362	0.362	0.357	0.343
Vegetables, no potatoes (cups)	0.380	0.359	0.366	0.340	0.331
Added sugar (tsp)	**0.361**				
Glucose index	0.371	0.360	0.354	0.349	0.341
Glucose load	0.385	0.353	0.363	0.348	0.335
Average energy expended (3day PAR)	0.384	0.368	0.359	0.344	0.341
Average vigorous PA blocks (3day PAR)	0.380	0.362	0.371	0.350	0.341
Average moderate-vigorous PA blocks (3day PAR)	0.378	0.365	0.359	0.351	0.342
Average sedentary PA blocks (3day PAR)	0.381	0.358	0.360	0.346	0.336
Weekly average sedentary PA	0.377	0.371	0.372	0.353	0.351
Weekly average moderate-vigorous PA	0.374	0.358	0.364	0.349	0.352
Weekly average vigorous PA	0.374	0.359	0.360	0.347	0.345

Figure 1. Study diagram for the Prevention (Prev) and Intensive (Int) Action PAC samples.

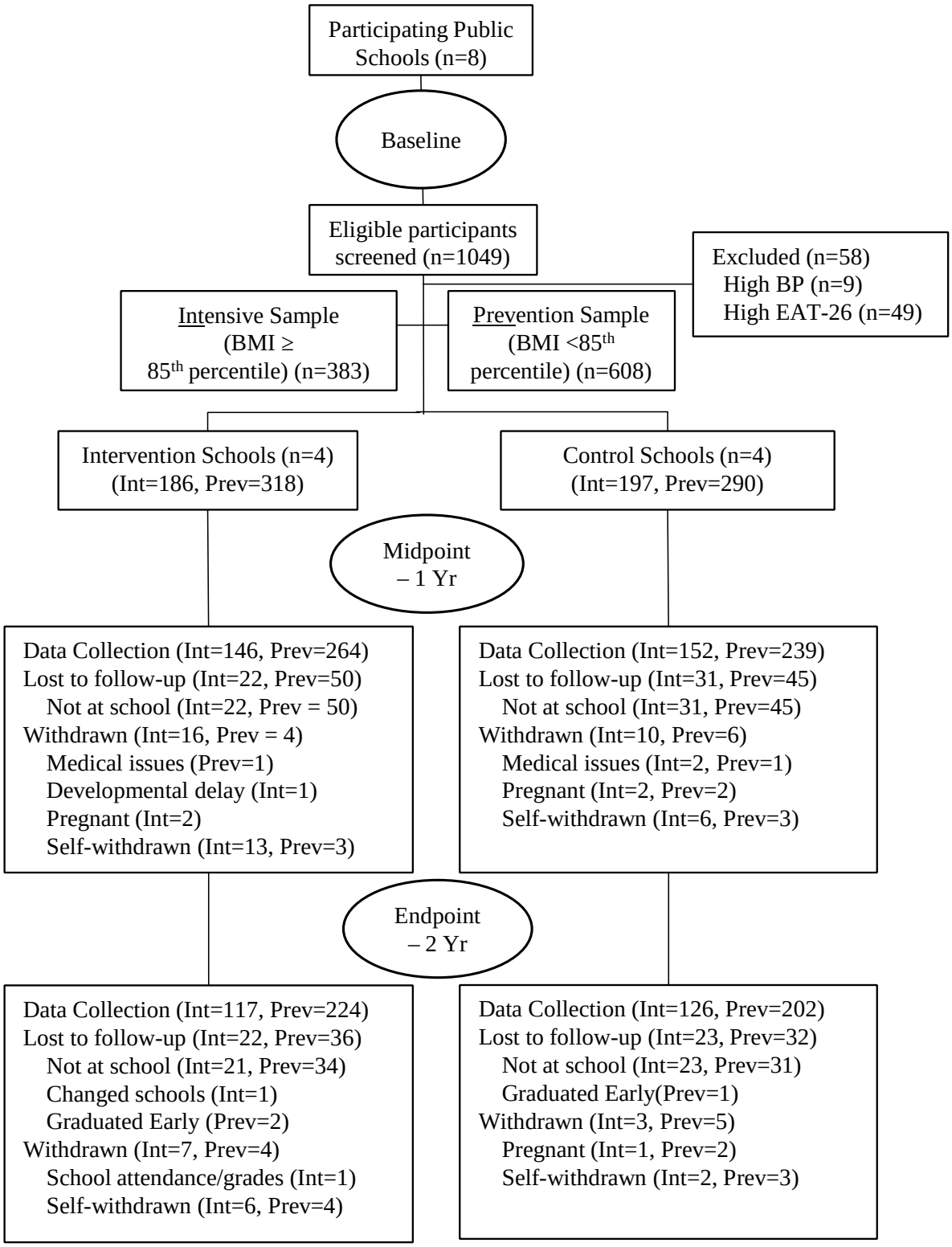

Figure 2. *Distribution of the predicted effects of the Action PAC preventive intervention versus control on BMI-z scores.*