Title: *BMi2+Population Effects of MI on Pediatric Obesity in Primary Care*

NCT03177148

Date: October 15, 2021

10/15/2021

BMI2+ Analysis Plan

Methods and Procedure

Randomization, concealment, and blinding

Randomization occurred at the practice level by a statistician blinded to practice identity. Practices were stratified based on estimates of racial and ethnic composition and size of their patient population, and the nearest neighbors of that sort became a pair. Practices were then randomized within pairs to either intervention or usual care groups. Since the number of practices was small, it was necessary to restrict matching to the variables deemed to have the biggest potential impact on BMI. Other variables will be considered as covariates in multivariable models.

Study cohort

To obtain the study cohort, we will first select control matches for each child in the intervention group from eligible children in the usual care practices. Potential characteristics used for matching include age (birth date), race/ethnicity, gender and, insurance and baseline BMI. If sufficient numbers of eligible control subjects are available, matches will be selected preferentially within the practice pairs established during randomization. In this approach an enrolled child in an intervention practice will be matched preferentially to eligible children in the corresponding control practice from the pair-wise randomization of practices. The enrollment date for each intervention subject will be assigned as a counterfactual "enrollment date" to their matched controls. We will exclude subjects without any follow-up BMI data.

Outcome measure

The main outcome is the % distance from the sex-age-specific 95% percentile over approximately 2 years. This metric recently has been shown to be a more reliable measure of adiposity than BMI z-score.(REF) In addition, this metric does not have an upper limit and can be used to assess BMI in all children, even those with high BMI values. Furthermore, BMI z-score correlates poorly with adiposity measures such as circumferences, triceps skinfold, and fat mass determined by dual-energy x-ray absorptiometry (DXA).37 We will investigate the BMI measure on both the linear scale (for interpretable effect measures) and log-transform scales (to meet required model/distribution assumptions). We will report the analysis results using both measures, one as the primary analysis and the other the sensitivity analysis. At the conclusion of the trial, BMI values for eligible children will be extracted from the PCC EHR and cleaned for implausible values as described above (see Methods and Analysis: Child Eligibility).31

Statistical Analysis

Once we have matched cohorts, we will conduct preliminary analysis to evaluate whether the process of randomization and matching have produced balanced distributions of children between intervention and usual care group. We will compare the balances of baseline characteristics of these children. If the distributions of certain characteristics are noticeably unbalanced between two arms, we will control for these characteristics in the regression during hypothesis testing.

For hypothesis testing, in our primary aim, we will compare the outcome of interest between children in the intervention group with their matched controls in usual care practices. The participants include those children of parents who were enrolled in the study. we will use mixed effect models to account for clustering of repeated BMI measures within children. The model accounts for clustering and considers different measures occurring at different times and at different frequencies. In the model, we will adjust for two time measures: calendar time and the elapsed intervention time (i.e. time since enrollment). A significant group effect indicates a significant difference between the two study arms. In addition, we will also fit a model including the interaction between time and intervention status variables. A significant interaction effect indicates the rate of BMI change differs between intervention and control groups. We will use the model to compare the predicted mean outcome between two groups at year 2 during follow up.

For our secondary aims, we will conduct similar analyses as follows: (1) compare the outcome of interest among children of participating Ped/NPs who actively participated in the intervention (defined as receipt of at least 50% of the total MI counseling sessions from Ped/NPs and RDs) versus all eligible children in usual care practices (analysis of completers per protocol) and (2) all eligible children of participating Ped/NPs in intervention practices, even if the child's parent did not enroll in the study, versus all eligible children in usual care practices (population-level analysis). For all our analyses, we excluded children from practices without any participating clinicians. In addition to defining active parent engagement in the intervention as receipt of at least 50% of MI counseling sessions, we will also examine whether there is a linear dose-response relationship between the number of MI counseling sessions received and study outcomes. Additionally, we will conduct descriptive statistics on the questionnaire data obtained during follow up for subjects in the intervention group.

Sample Size and Power Calculation

The study was initially powered to detect an effect of 0.10 BMI z-score units (the original metric used in the grant application) between intervention and usual care groups at 2-year follow-up, with an assumed standard deviation of 0.40, power of 0.80, and 2-tailed alpha of 0.05. This equates to a standardized effect of 0.25 (0.10/0.40). To convert this same effect size to our revised outcome, % from median BMI for age and sex, we used data from Freedman et al.35 Assuming our sample (all of whom are above the 85th percentile) will average around 30%-40% above the median (which equates to the 97th percentile), with a standard deviation of 20-25, this equates to a change of approximately 5-6 percentage-from-median units. We will log transform our primary outcome variable as needed.
Sample size estimates account for practice-level clustering,40 assuming a practice-level intraclass correlation coefficient between 0.001 and 0.03.

Based on these assumptions, we required 7 practices per study group (14 total, although 2 additional practices were recruited in each study group to account for possible practice attrition) and an average of about 35 enrolled parents per intervention practice (target n=316).