

The START study: An evaluation to study the impact of a natural experiment in high school start times on adolescent weight and related behaviors

Rachel Widome¹ widome@umn.edu
Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN, USA

Kyla L. Wahlstrom wahls001@umn.edu
Department of Organizational Leadership, Policy and Development, College of Education and Human Development, University of Minnesota, Minneapolis, MN USA

Melissa N Laska nels5024@umn.edu
Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN, USA

Darin J Erickson erick232@umn.edu
Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN, USA

Aaron Berger berge314@umn.edu
Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN, USA

Conrad Iber iberx001@umn.edu
Division of Pulmonary, Allergy, Critical Care and Sleep Medicine, University of Minnesota Medical School, Minneapolis, MN, USA

Gudrun Kilian¹ kilian@umn.edu
Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN, USA

¹ Corresponding author.

Abstract

Background: Research has shown that early high school start times, which are asynchronous with adolescent biology, are one of the most significant obstacles to youth being able to get sufficient sleep. Given that adolescence is a critical period that sets the stage for long-term obesity risk behavior patterns, there is a need to understand the obesity-related implications of increased sleep as a result of intervention and policy changes.

Methods: We evaluated a community-based natural experiment in school start time policy modification when several Minneapolis-St. Paul, MN metro area school districts shifted to later school start times in Fall 2016. We collected data on student weight and related risks (via paper survey, objective weight and height measurement, dietary recall, and sleep actigraphy) before and after two districts (two high schools) shifted their start times later and in a comparison district (three high schools) which kept their start times early (7:30am)

through the course of the study. Our specific aims were: 1) Determine how a shift to a later high school start time relates to objectively measured weight change over time. 2) Identify the relationship between school start times and obesity-related behaviors over time.

At baseline we had 2,133 returned surveys (93% participation) and 2,037 (86% participation) objective height/weight measurements from 9th grade students (class of 2019) in the five schools. The sample was 87.7% white, 12.8% reported qualifying for free/reduced price lunch (a measure of lower socio-economic status), and the mean age was 15.2 (SD=0.35) years.

Discussion: The products of this research will clarify causal connections between sleep and obesity among adolescents as well as provide evidence for whether a school start time policy can minimize unhealthy weight gain.

Keywords: Natural experiment, school start time, sleep, obesity

1. Background

Although the National Sleep Foundation recommends 8.5 to 9.25 hours of sleep per night for adolescents (National Sleep Foundation 2000), more than half of US 16-year-olds regularly get less than seven hours (Keyes et al., 2015); further, disadvantaged youth are at even greater risk for short sleep duration (Keyes et al., 2015). “Sleep debt” has consequences that extend far beyond hampering day-to-day functioning. It is deleterious to multiple areas of physical and mental health (Institute of Medicine (US) Committee on Sleep Medicine and Research 2006). In the context of the obesity epidemic, which has had a far-reaching public health impact, sleep insufficiency is emerging as a potentially important and novel risk factor in the development of excess weight gain and weight-related behaviors in children (Owens et al., 2014). In fact, one meta-analysis concluded that children and adolescents who do not get sufficient sleep have doubled odds of overweight and obesity over time (Fatima et al., 2015). However, there have been inconsistencies in findings linking sleep and obesity, and this literature has been primarily built around observational and cross-sectional research. Thus, whether associations are truly causal has not been established.

While a variety of forces such as caffeine intake, screen time, homework expectations, and other factors can be implicated in stymying American teenagers' sleep, research has shown that high school start times, which are asynchronous with adolescent biology, are one of the most significant obstacles to youth being able to net sufficient sleep (Group and Committee on Adolescence 2014). When children begin puberty, circadian rhythms shift such that melatonin is secreted later in the evening making it difficult to fall asleep before 11pm. This neurobiological change means that adolescents who attend "early-start" (prior to 8:30am) high schools, even those who strive for an early bedtime, will struggle to get sufficient sleep. Several evaluations of natural experiments have suggested that later school start times allow adolescents a greater opportunity to accumulate adequate sleep and that students reap academic, mental health, and safety benefits when their schools start later (Edward et al., 2008; Marx et al., 2017; Minges and Redeker, 2016; Wahlstrom et al., 2014; Wheaton et al., 2015, Group and Committee on Adolescence 2014). With this evidence, multiple organizations, including the American Academy of Pediatrics, the American Medical Associations, and the American Academy of Sleep Medicine, that focus on health have recommended that high schools start at 8:30am or later (Group and Committee on Adolescence 2014; Start School Later n.d.). However this is in sharp contrast to reality for most US adolescents, where less than 15% of US high schools start at 8:30am or later and 46% start very early, at 8am or before (Anon n.d.; Taie and Goldring n.d.; Wheaton et al. 2015), which appears to most severely curtail sleep (Taie et al., 2017).

To move forward and advance the evidence basis for how to optimize the timing of school schedules, there are methodological weaknesses in the previous research that must be addressed. School start time is an inherently difficult exposure to study for a variety of reasons. These include the inability to randomly assign start times to schools in the US, the challenges of collecting baseline data prior to a start time change (which, in practice, are often deployed soon after policy decisions are made), and the difficulty in addressing both internal and external validity when a relatively limited number of schools participate in any given study. Further, previous studies have not used objective measures of outcomes and typically have limited follow-up time. Given this, multiple recent reviews have called for additional research that could overcome some of these challenges to internal and external validity (Marx et al., 2017; Minges and Redeker, 2016; Wheaton et al., 2016).

Acknowledging that: 1) so few teens are attaining sufficient sleep, likely in a large part due to their early school starts, 2) observational research suggests a link between sleep deficiency and obesity, and 3) adolescence is a critical age that sets the stage for long-term obesity risk behavior patterns, there is both a need and opportunity to understand the obesity-related implications of increased sleep as a result of intervention and policy changes, such as those to delay high school start times. To address these issues, we designed the START study, an evaluation of a natural experiment in high school start time modification. The overarching objective of the START study was to determine how high school start times influence obesity and related behaviors. We aimed to design a study that could fill several of the methodological gaps of previous work. We hypothesized that the benefits of later start times, an established effective sleep intervention, may extend further – to promoting weight-related health. Our proposed specific aims were to:

- 1) Determine how a shift to a later high school start time relates to objectively measured weight change over time. Our hypotheses were that students experiencing later start times will report less of a longitudinal increase in body mass index (BMI) over the study period, due to increased sleep and that this effect will be evident across gender and socioeconomic status (SES).
- 2) Identify relationships between school start times and obesity-related behaviors over time. Our hypotheses, as supported by previous sleep-related research, are that later start times will be associated with healthier eating behaviors (like reduced breakfast skipping and sugary beverage intake), greater physical activity, and less sedentary behavior, and that this association will be evident across gender and SES.

Additionally we specified an exploratory aim, to assess barriers and facilitators associated with implementing later high school start times. Our hypothesis is that both concerns and actualities about how later start times impact family, sports, and work schedules could make the policy seem more or less palatable to stakeholders.

Here we will describe START methods in depth as well as describe the profile of the cohort we enrolled at the baseline (Spring 2016) measurement.

2. Methods/Design

Study overview

To test our hypotheses, we designed the longitudinal START study to examine weight-related outcomes of a natural experiment in school start time policy. We recruited two Minneapolis, MN metro area school districts (each with a single high school) which had made the decision to shift the start times of their high schools from 7:30am and 7:45 to 8:20am and 8:50am, respectively, when the 2016/17 academic year began. Additionally, we recruited one comparison school district with three high schools all with a 7:30am start time and no plans to alter their schedules in the coming years. In all five high schools, the START study would collect data on student weight and related risks in a cohort of students who were in the 9th grade in 2015/16, following them over three school years in both the policy-change and comparison schools. Data collection for START included one wave of data collection before as baseline – Spring 2016, when all schools started at 7:30am, and then in two waves after (follow-ups 1 and 2 – Spring 2017 and Spring 2018) when the start time had changed for the schools that were shifting schedules later. Data would be collected at the same three timepoint in the comparison district. All students in the five schools who were in 9th grade in 2015/16 were invited to participate in the survey and objective BMI measurement. A subsample of these 9th grade students were also invited to complete

sleep logs, dietary recall measures, and wear wrist actigraphs.

Time sensitivity

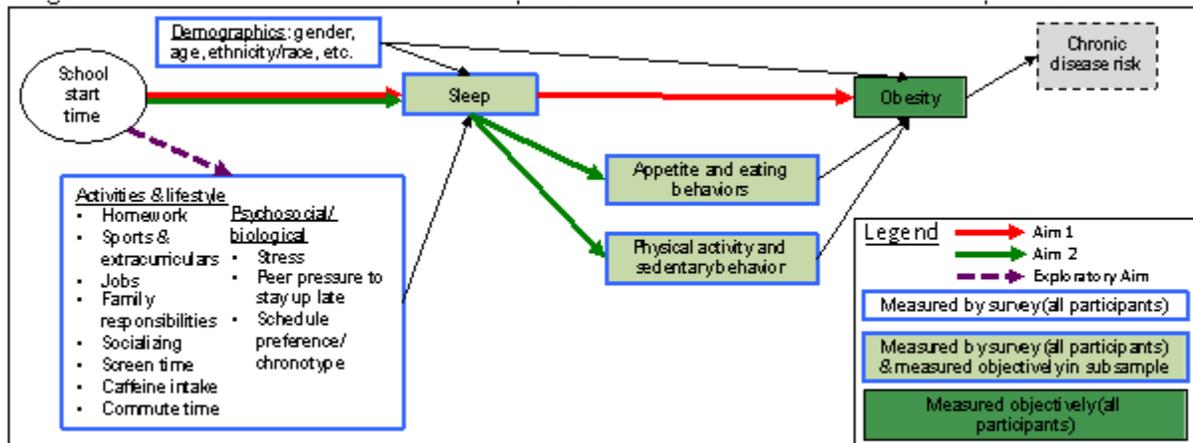
In order to accomplish this research, this study had to be launched in an uncommonly expeditious manner. Typically there is a period of uncertainty where schools discuss and debate the prospect of shifting high schools for one or more years before a solid decision is made to move forward with the change. After a decision is finalized, the schedule change is typically implemented 6-12 months in the future, at the start of the following school year. Thus, these policy change processes can be broken into two phases: a policy discussion phase of 1-3 years and a 6-12 month phase between policy decision and enactment. This timeline offers the advantage of being able to identify a comparison district(s), for whom lack of policy discussion indicates that they are at least several years out from any potential change. The disadvantage is that the 6-12 month window after the policy is decided on does not provide sufficient time to secure research or evaluation funding through traditional funding cycles after a school that is certain to change has been identified. In order to collect measurements on the adolescents, ahead of the shifts to a later start time in the intervention districts, we applied for and successfully received funding from the NIH's "Time-Sensitive Obesity Policy and Program Evaluation (R01)" Program Announcement (PAR-12-257). Without having access to this unique mechanism, this large-scale study of the impact of school start time changes on weight would not have been possible.

Theoretical framework

The theory that guides this proposal is Fundamental Cause Theory (Phelan et al., 2010), which explicates that there are resources, such as prestige, financial capital, and social capital that protect general health. We posit that sleep is also one of these resources. According to Link and Phelan's definition, a "fundamental cause" of disease involves a lack of access to a resource, which could have been used to avoid or mitigate disease. A resource deficit, driven by a fundamental cause, will contribute to multiple incursions on health, irrespective of the mechanism of the particular health threat. A classic example is how financial strain might put someone at risk for both mental health issues and diabetes even though these health outcomes have very different natural histories. A force, such a school schedule, that is asynchronous with adolescent neurobiology and thus restricts access to sleep follows this model as lack of sleep will influence multiple, disparate health outcomes. Sleep debt fits this model as it is connected to a variety of disease states including cardiovascular disease, injury, and poor mental health, among others (Institute of Medicine (US) Committee on Sleep Medicine and Research 2006).

According to this model, anyone with a sleep resource deficit will be more vulnerable to adverse outcomes. However, those who face disadvantage and additionally have lesser access to other resources (i.e. social capital, information, money), may be set back even further by lack of sleep. This is why a school schedule that is more conducive to sleep, while it stands to benefit all students, may be especially important for students who have disadvantage in other areas. Drawing from Fundamental Cause Theory, Figure 1 depicts the proposed causal relationship between school start time and obesity, as well as the framework for mea-

Figure 1. Causal model for the relationship between school start time and obesity.



measurements for the START Study.

Study setting and population

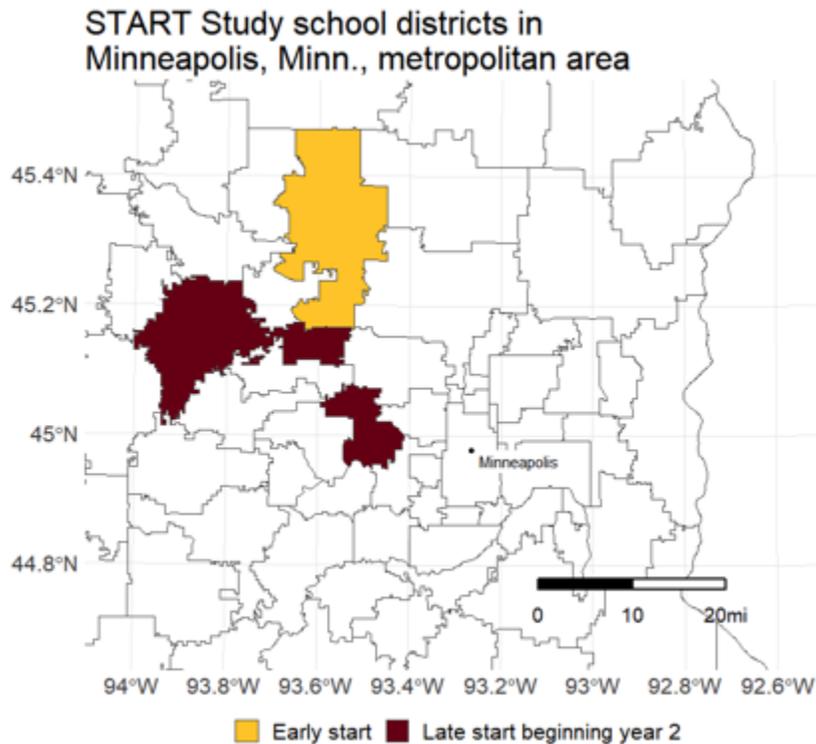
START compared school districts where high school students start at 7:30am or 7:45am in 2015/16 but would shift to an 8:30am or later start time in 2016/17 and beyond, to a school district where high school students start early (7:30am) in 2015/16 with no expectation of change. An important advantage is all districts are in the Twin Cities, MN metro area and thus similarly positioned in the Central Time Zone (See Figure 2 for a map of districts). The means that students in all schools experienced the same day lengths and sunrise times.

-Start time change school districts:

- 1) The **Wayzata School District** encompasses eight suburban communities within the Minneapolis-St. Paul, MN metro area and is located approximately 18 miles northeast of downtown Minneapolis. Wayzata high school serves the entire school district, includes grades 9-12 and has an enrollment of just over 3,200 students (Anon n.d.). In 2017, 11.3% of students qualified for free or reduced price lunch and 28.5% of the student body was racial/ethnic minority (Anon n.d.).
- 2) The **Buffalo School District** is an exurban school district located about 37 miles northeast of downtown Minneapolis and includes grades 9-12 in a high school serving the entire district. In 2017 there were approximately 1,859 students enrolled, 9% of whom were racial ethnic minorities and 19% were eligible for free and reduced lunch (Anon n.d.).

-Comparison school district: As we sought comparison schools, we prioritized finding communities that were as close a match as possible, demographically, culturally, and geographically to the policy intervention districts. These considerations led us to select the **Elk River School District School**, which is located approximately 30 miles northwest of downtown Minneapolis. Elk River has three high schools, 1) Elk River High School, 2) Rogers High School, and 3) Zimmerman High School which serve this exurban/suburban district. In

Figure 2. Map of START study school districts.



2017, in the Elk River School District, the proportion eligible for free and reduced lunch ranged from 12%-20% and the proportion of students were racial/ethnic minorities ranged from 8-11% (Anon n.d.). In Fall of 2015 there were approximately 3,570 students attending the three Elk River district high schools.

Data collection procedures and measures for survey and anthropometry

Informed consent. Prior to conducting the first in-school measurements in each year of the study, we provided the school districts with information sheets to be distributed to each student's household by e-mail. This information summarized the study, informed parents/guardians of the study's upcoming survey that will be administered in-school, and explained that data collectors will be coming to the school to take height and weight measurements. The information sheet included elements of consent (that having measurements taken is voluntary, data will be kept confidential, etc.). Parents/guardians of students who did not wish for their child to participate could call or e-mail the research team or school office prior to the data collection day to opt their child out of research activities. On data collection day the research was explained to students. Our participant assent process communicated that students were free to refuse participation in any element of data collection with no consequence to them and it gave students the opportunity to ask questions about the study. Both the University of Minnesota Institutional Review Board and participating school districts approved a "waiver of documentation of informed consent" procedure for

the START study.

Survey. All students in 9th grades in all five schools in Spring 2016 (baseline) were invited to complete the survey at baseline (Spring 2016) and again in Spring 2017 and 2018 (follow-ups 1 and 2). These paper surveys were designed to be completed in approximately 15 minutes. Table 1 summarizes the survey measures which were selected based on study aims. Most items were derived from existing validated measures and surveys that have been used in large scholarly studies of adolescents including the Teen Sleep Habits Survey (Anon.n.d., 015a; Wahlstrom et al., 2014), and Project EAT (Eating Among Teens) (Neumark-Sztainer et al., 2002). The use of measures that had been previously fielded and/or validated was necessary as we were not able to pilot the START survey given the time-sensitive nature of this project. However, during our survey development period, research team members, expert colleagues who were outside of the study but in the areas of sleep, weight, and health behavior, and adolescent consultants reviewed drafts of the survey and minor modifications were made.

Table 1. Survey measures.

Measures	Brief description
Weight status*	Height and weight asked of all participants.
Demographics, individual behaviors, and environment	Age, race/ethnicity, gender, depression, stress and ability to manage stress
Daily schedule	Extracurriculars & jobs, homework, family care responsibilities
Sleep	Adaptation of Wahlstrom et. al's Teen Sleep Habits Survey (19,20) which includes items on number of hours slept on weekends and weekdays, bed and wake up times, reasons for going to bed and wake when they do, sleep quality and daytime sleepiness, negative sleep-related events, sleep distractions.
Eating patterns	Items on frequency of eating breakfast, fast food, fruits and vegetables, sugar sweetened beverages, caffeine, sweets/desserts (adapted from NCI's Dietary Screener Questionnaire [DSQ](33) and Dietary History Questionnaire [DHQ](34)), family meal frequency, dieting and other weight control behaviors, binge eating (Adapted from Project EAT (35)).
Physical activity (PA) and sedentary behaviors	Weekly hours of strenuous, moderate, and mild physical activity, PA related to work, transportation, household chores, and leisure-time and sedentary behavior, leisure screen time (adapted from PDPAR (36), Godin-Shepard (37), PAQ-A (38), and IPAQ (39))

Objectively measured height and weight. At four of the five schools (Elk River, Rogers, Zimmerman, and Buffalo), all students in the 9th grade were eligible to have objective height and weight measurement at all three waves of data collection. In one school, Wayzata, 296 of the 852 ninth grade students were not available to be measured as this school requested that we conduct measurements during physical education (PE) class be-

cause these students had already completed their 9th grade PE class in the Fall semester and the START study was launched in the Spring term. In subsequent waves, all students in the class of 2019 at Wayzata were invited for height and weight measurement. Our protocol for taking anthropometric measurements was adapted from Project EAT (Project Eating and Activity among Teens), which has successfully collected high-quality height and weight data from thousands of adolescents in school settings (Neumark-Sztainer et al., 2002).

On measurements days, height and weight measurement stations were set up in the schools in a private area (either a separate room or behind a portable screen) such that the scale digital display was not visible to any non-study personnel. As students entered the study area data collectors introduced themselves and explained to students that having their height and weight measured for the study is voluntary and for the purposes of a University of Minnesota research study. Students could decline measurement within the private area such that it would not be obvious to other students, teachers, or non-study staff who chose or refused participation. Data collectors were trained, certified, and re-certified annually based on height and weight measuring protocols that had been successfully used for other studies in school settings. On each measurement day validated readings by having select participants measured by two different data collectors to assure reliability and validity.

Using a Tanita BWB-800, weight was measured twice, recorded to the nearest 0.1 kg, and then averaged for each participant. If the two measurements differed by more than 0.2 kg, a third measurement was taken and the final weight would be the average of the three readings. Scales were placed on firm, flat surfaces in the measurement station and calibrated using a 5kg calibration weight. Height was measured twice using a Seca Model 217 stadiometer, to the nearest 0.5 cm and then the two measurements were averaged. Similar to weight, measurements differed by more than 0.5 cm a third measurement was taken and all three readings would be averaged. Participants were asked to remove shoes, coats or heavy outerwear, and heavy items from pockets prior to measurement. In the case of a student who would like to participate but has physical injuries, disability, pregnancy, or other issues that prevent them from being measured according to protocol (such as discomfort with removing headwear), we made modifications to the protocol so that the student can participate in a way that was safe and comfortable for them, and noted the protocol deviation in our data collection form. Height, weight, any notes about protocol modifications, and the students' IDs (to link with survey data) were recorded on data collection forms. Students could request their own height or weight numbers if this was their preference. Participants received a small (\$1-2 value) thank you gift for participation.

Data collection procedures and measures for substudy (actigraphy and diet)

Objective measurement of sleep and diet substudy. At Baseline we randomly selected a subsample of survey participants to be invited to participate the objective sleep and diet measurement substudy. For the substudy, the number of participants selected at each school was proportional to the school's 9th grade class size in order to ensure we would have an adequate of students from our smallest school for analysis. We collected data from 288 participants at Baseline and then re-contacted Baseline substudy participants for Follow-ups 1 and 2. In waves subsequent to Baseline, if substudy participants were no

longer attending a participating school, they were not eligible for follow-up substudy data collection. The substudy had time-alignment with the survey data collection (in the Spring of 2016, 2017, and 2018) in order for substudy measures to have maximum possible utility for survey validation. Our substudy recruitment procedures were as follows: An information letter (and follow-ups) about the substudy was sent to the randomly selected students' parents or guardians via an e-mail from the school, with instructions to phone or e-mail study to consent to their child's participation. Youth also had the opportunity to assent when substudy materials were distributed.

After parents or guardians gave consent for their child's participation, data collectors met with small groups of prospective substudy participants at their schools to get assent for participation from the youth and distribute actigraphs (wGT3X-BT Monitor, Actigraph, Pensacola, FL, USA), an unobtrusive, wristwatch-like device that collects objective data on movement, light, time spent sleeping vs. awake, and sleeping patterns. Actigraphy has been shown to correlate well with polysomnography (PSG) (Ancoli-Israel et al. 2003), which is the gold standard of objective sleep measurement, but actigraphy is more practical, cost-effective, and less intrusive. We instructed participants to wear the actigraphs for seven consecutive days on the non-dominant arm, removing them only for bathing, swimming, or other activities where the arm would be fully immersed in water (the device could be worn for handwashing and showering). Sleep parameters derived from activity will be scored by the Sleep Reading Center at Brigham and Women's Hospital (PI: S. Redline). Data from the actigraphs would be used to calculate measures including sleep latency onset (time from laying down to sleep onset), total minutes asleep, sleep efficiency (percent of time in bed sleeping), and number of awakenings after sleep onset. During the week that the participants wore the actigraph they were also asked to complete a sleep log where they recorded when they got in and out of bed, whether the device was removed for any periods, and whether they took any naps. This information was used to cross-validate the actigraphy data. A week after distributing the materials, data collectors visited schools to retrieve the actigraphs and sleep logs and distribute incentives.

The substudy students were also invited to complete a web-based, self-administered 24-hour food recall, the ASA24-2016 which is a standardized, comprehensive measure of dietary intake. The ASA24-2016 online system guides the participant through recalling and recording all they had consumed in the preceding 24 hours. Visual cues in the system guide subjects to report on each eating occasion, timing of eating, what was consumed, and portion size. ASA24-2016 was designed to follow the USDA Automated Multiple-Pass Method (AMPM) which is has been validated against biomarkers of energy intake (Jennifer et al., 2012; Alanna J et al., 2008); further, yet-unpublished, data shows close agreement of the newer ASA24-2016 with interviewer-administered 24-hour recalls (Applied Research: Cancer Control and Population Sciences n.d.:24). These measures can be used in analyses as outcomes and will be used for validate survey diet measures. Substudy compensation was as follows: \$35 (gift card) for returning the actigraph, \$20 for returning the sleep log, and \$30 for completing the ASA24-2016 – for a total possible compensation of \$85.

Analysis. The study design is a 2 (school start time conditions: early or late) x 3 (Timepoints: Spring 2016, 2017, Spring 2018) natural experimental evaluation. The primary hypothesis is that, compared to those who start school early (at 7:30am or 7:45am), those whose schools shifted to a later start time by Follow-up 1, will show improvements in obesity-related behaviors and weight change over time as measured by survey items and objectively measured BMI (calculated as weight in kg / height in m^2). We will use a random effects repeated measures model to test this hypothesis.

The target estimand is the differential change in BMI and other weight-related behaviors in schools that experienced a policy that shifted their start time later, compared to schools that remained at an early start time throughout the observation period (9th through 11th grade). Specifically,

$$Y_{ti} = b_0 + b_1TIMEPOINT_t + b_2COND_i + b_3TxC + B_kC + \varepsilon_{ti}$$

where Y_{ti} = the outcome for each individual i (1 to N) at each timepoint t (1 to 3), b_0 is the intercept, b_1 is the regression coefficient associated with the main effect of time, b_2 is the regression coefficient associated with the main effect of condition (school start time: early or late), and b_3 is the regression coefficient associated with the interaction between time and condition. The model includes a matrix C representing covariates, and ε_{ti} represents individual residuals. The regression coefficient for the interaction term (b_3) is the estimate of the treatment effect under the common trend assumption.

Our method (a Difference in Difference approach) (Dimick and Ryan, 2014; Wing et al., 2018) for estimating the causal effect of this policy relies on the assumption that in the absence of the policy change, the group that received the policy change would have had a parallel trend in its outcomes over time as the comparison group that did not receive the policy change. While these assumptions are strong, we feel they are reasonable given the demographic and geographic similarities in our school districts. We are somewhat protected against an exogenous factor influencing schools differentially over time as there were no interventions or policies with solid evidence bases for impacting weight change that were planned to be implemented in either condition. Our design does not allow for us to test equivalence of pre-intervention trends since there was only one wave of data collection prior to the policy change.

Specific Aim 1 focused on adolescent weight change over time. We have chosen to use actual BMI change rather than BMI Z-score change as our main outcome measure as recommended by Berkey and Colditz (Berkey and Colditz, 2007) as best practice for longitudinal studies of growing adolescents. (BMI Z-scores are based on CDC growth charts and derived from cross-sectional data, and thus are thought to potentially misrepresent the growth patterns of children measured repeatedly (Berkey and Colditz, 2007).) In our first aim’s analyses, we will test our hypothesis by regressing each BMI on school start time, timepoint, and the school start time x timepoint interaction, as well as a number of potential covariates that could be confounders of the exposure – outcome relationship. For all primary analyses (both Specific Aims 1 and 2), we will include the following prespecified confounders: biological

sex (male / female / prefer not to respond), free and reduced-price lunch eligibility (yes / no / don't know), Hispanic ethnicity (yes / no), race (Native American / Asian / Hawaiian or Pacific Islander / Black / White / Multiple / unknown or not reported), each parent or guardian's education (did not finish high school, finished high school/GED, some college, finished college, advanced degree, don't know, do not have a second parent or guardian). We will test the primary hypothesis using the school start time x timepoint interaction. A significant two-degree of freedom interaction indicates differences between students in the districts that switch to later school start time versus the district that does not. Specific Aim 2 will be tested similarly with weight related behaviors as the outcomes.

Random effects repeated measures models for continuous outcomes (e.g., BMI) will be estimated using the MIXED procedure (in SAS), which can estimate a large number of models, including general linear mixed models designed to analyze repeated measures (Fitzmaurice et al., 2011). Categorical outcomes (such as in the case where the outcome is obesity category) will be modeled using PROC GLIMMIX. There may be some correlation of observations within the schools (given that there will be two intervention and three comparison schools). However, based on intra-class correlations for BMI that we calculated using 10th grader data from Project EAT (Neumark-Sztainer et al., 2002), which sampled youth from 36 schools in the Minneapolis/Saint Paul metro area (ICC = 0.010), variance inflation for this within-school clustering will be minimal (Project EAT ICC previously unpublished). However, we planned to account for any unforeseen correlation due to nesting of students within schools by including a random intercept for school in models. We will use a repeated measures analysis to account for the longitudinal within-subject correlation. Specifically, this model relaxes the constraint of residual independence, allowing residuals to correlate to account for the same students being measured over time. We expected that sleep data missing due to drop-in and drop-out will follow a missing-at-random (MAR) pattern. In this scenario the data are efficiently handled with maximum likelihood estimation in linear mixed models, which we will use in all analyses (Allison, 2012). This study provides an opportunity to test this assumption as we can compare self-reported BMI among those who have an objective BMI versus those where objective BMI is missing.

To cross-check survey measures, we will examine correlations between select self-reported sleep items and objectively-measured sleep characteristics (actigraphy data) overall and within subgroups (i.e. gender, SES). We will also examine correlations between 24-hour food recall (ASA24-Kids-2014) and select survey items included in the survey from the NHIS dietary screener to assess validity of those screener items in our population. If needed, a correction factor for survey items can be derived. We will also conduct sub-study analyses relevant to Specific Aim 2, that utilize data from the substudy (for instance sleep latency or nutritional content of breakfasts) that only exists for substudy participants.

Sample size and power (Aims 1 and 2). Because this is a study of a naturally occurring experiment in select school districts, the number of potential participants has an upper limit. We estimated a conservative survey response (70%) at baseline and expected to be able to obtain objective heights and weights on 80% of baseline survey responders. We estimated we could retain 80% of baseline responders at Follow-up 1 and 90% of Follow-up

1 responders at Follow-up 2. With these rough assumptions and given enrollments at the schools we expected would participate, we anticipated having 1,180 students with survey and objective anthropometric data at all three time points. As we illustrate below, this yields an ample sample size for both longitudinal and cross-sectional analyses and we will be able to utilize all data collected, even if a participant only participated at only one or two timepoints, at least in select analyses.

Using means and variances from studies measuring obesity and related outcomes in these or similar schools (Hibbs et al., 2013; Neumark-Sztainer et al., 2002) and setting $\alpha=0.05$ (two-tailed), Table 2 displays power to detect differences in various scenarios, including subgroups analyses and other situations where we are not able to utilize the full n. For instance, Scenario 1 shows that we have 0.97 power to detect a situation where adolescents in the comparison district gain 0.5 units in BMI more than adolescents in the policy intervention districts between 10th and 12th grade. As shown in Table 3, the survey and anthropometric portions have ample power to detect meaningful differences, even for subgroup (gender, SES) analyses or if attrition was to be higher than expected.

Table 2. Power analysis for detecting differences in various scenarios.

Scenarios - Ability to detect:	Est. mean (SD)	Policy n	Comparison n	Power
1) 0.5 BMI unit longitudinal change difference over the course of 3 time points, with projected sample	22.3 BMI units (4.5)*	658	717	0.97
2) 0.5 BMI unit longitudinal change difference over 3 time points if final sample is 30% smaller due to unexpected attrition	22.3 BMI units (4.5)*	461	502	0.90
3) 0.5 BMI unit longitudinal change difference over 3 time points if the larger intervention s fails to implement policy	22.3 BMI units (4.5)*	197	1178	0.77
4) 0.5 BMI unit longitudinal change difference over 3 time points among girls	22.3 BMI units (4.5)*	329	358	0.78
5) 1.0 BMI unit longitudinal change difference among low-income student population over the course of 3 time points	22.3 BMI units (4.5)*	164	179	0.97
6) 0.2 breakfasts per week longitudinal change difference over 3 time points.	3.2 per week (1.5)*	658	717	0.96
7) 0.15 sugar sweetened beverages per day longitudinal change difference over 3 time points	1.4 per day (1.1)*	658	717	0.96
8) 0.3 breakfasts per week longitudinal change difference over 3 time points among low-income participants	3.2 per week (1.5)*	164	179	0.79
9) 0.25 sugar sweetened beverages per day longitudinal change difference over 3 time points, among low income	1.4 per day (1.1)*	93	143	0.88

*Mean overall values and standard deviations from Project EAT study, unpublished

^Mean values and standard deviations from Redline et al. (31)

Qualitative analysis. Despite extensive evidence that later start times have been shown to have academic benefits for students and contribute to their mental wellbeing, the vast majority of high schools in the US start prior to 8:30am (Anon.n.d., 015a; Taie et al., 2017; Wheaton et al., 2015) . To explore barriers and facilitators to school schedule policy change, we will conduct in-depth interviews with stakeholders (principals, teachers, other school staff, parents, and students) in the policy intervention districts in Year 4 of the study, after the later start time has been in effect for the intervention schools for several years. Working with the intervention school districts, we will develop a list of potential participants that represent various facets of the school community to be invited to a one-on-one interview. The semi-structured interview guide will cover topics such as: (1) how have policy changes influenced childcare issues for younger children, athletics, family schedules, youth work for pay; (2) impressions of impact of later start time on school performance and on other aspects of health, such as mental health; (3) transportation; (4) policy fidelity (i.e., is the impact of a late start time attenuated by additional pre-school activity scheduling?), and (5) what potential benefits or drawbacks of the shift are most and least compelling given participants’ situations. We aim to conduct 15 interviews. Interviews will be audiorecorded and transcribed. NVivo software (QSR International, Melbourne, Australia) will be used for data analysis. We will use an iterative grounded theory method, the constant comparative process, to identify key themes. Coding by two analysts will be done on a subset of transcripts to assure reliability (Glaser et al., 1967).

Demographics at Baseline

Table 3 shows that at our baseline response proportion for survey was 93% and 86% for objective height and weight measurement. 28% of those invited in the substudy at baseline participated. The mean age of participants at baseline was 15.2 (SD = 0.35).

Table 3. Participation proportions for START study at Baseline (total and by school).

	Total	Buffalo	Elk River	Rogers	Wayzata*	Zimmerman
	%					
Student survey	93	93	89	93	96	92
Objective height & weight	86	55	83	87	95	78
Both survey and height/weight	71	68	80	82	60	89
Sub-study	28	25	37	35	25	18

*There were 852 students in the 9th grade at Wayzata High School eligible for surveying but 556 eligible for objective height/weight measurement due to the school’s physical education class schedule.

Our sample was predominantly white, roughly half female, and 13% reported qualifying for free/reduced lunch (Table 4). Approximately 30% reported sleeping less than 7 hours on school nights. There were differences in demographics between schools with Wayzata having the largest proportion of non-white students and Zimmerman having largest proportion of low SES students, measured by proportion of students reporting that no parent or guardian finished college and proportion reporting that they qualify for free or reduced priced lunch. Proportion of students who were categorized as obese based on objective height and weight

measurements ranged from 9.4% (Wayzata) to 16.5% (Buffalo). Table 5 shows differences between the survey and substudy samples. The substudy participants were less likely to be of a lower socioeconomic status.

Table 4. Demographics and sleep duration of START study survey responders at Baseline (total and by school).

	Total	Buffalo	Elk River	Rogers	Wayzata	Zimmerman	p-value ¹
	N (%)	N (%)					
Sex							0.12
Male	1066 (50.9)	213 (48.4)	174 (47.3)	185 (55.2)	412 (51.1)	82 (56.6)	
Female	1028 (49.1)	227 (51.6)	194 (52.7)	150 (44.8)	394 (48.9)	63 (43.5)	
Race/Ethnicity (check all that apply)							
White or Caucasian	1802 (87.7)	397 (92.5)	332 (91.2)	311 (94.5)	621 (78.8)	141 (96.6)	<.0001
Black, African, or African-American	118 (5.7)	17 (4.0)	17 (4.7)	12 (3.7)	71 (9.0)	1 (0.7)	<.0001
Hispanic or Latino/a	73 (3.5)	17 (3.9)	14 (3.8)	9 (2.7)	29 (3.6)	4 (2.7)	0.87
Asian	180 (8.8)	11 (2.6)	14 (3.9)	11 (3.3)	143 (18.2)	1 (0.7)	<.0001
American Indian or Alaskan Native	70 (3.4)	20 (4.7)	16 (4.4)	10 (3.0)	16 (2.0)	8 (5.5)	0.05
Native Hawaiian or other Pacific Islander	13 (0.6)	2 (0.5)	4 (1.1)	3 (0.9)	3 (0.4)	1 (0.7)	0.50
Something else	18 (0.9)	14 (3.2)	1 (0.3)	0 (0.0)	3 (0.4)	0 (0.0)	<.0001
Parent Education							<.0001
Finished college	1502 (79.5)	259 (69.8)	227 (72.3)	239 (79.9)	697 (90.5)	81 (59.1)	
High school or some college	388 (20.5)	112 (30.2)	87 (27.7)	60 (20.1)	73 (9.5)	56 (40.9)	
Qualify for free/reduced lunch							<.0001
Yes	258 (12.8)	74 (17.9)	41 (12.4)	36 (11.1)	79 (9.9)	28 (18.8)	
No	1383 (68.6)	249 (60.1)	224 (67.9)	203 (62.7)	612 (76.7)	95 (63.8)	
I don't know	374 (18.6)	91 (22.0)	65 (19.7)	85 (26.2)	107 (13.4)	26 (17.5)	
Sleep duration (in hours)							0.005
< 6.0	190 (9.1)	31 (7.2)	30 (8.3)	34 (10.1)	75 (9.3)	20 (13.7)	
6.0-6.75	408 (19.6)	58 (13.5)	88 (24.2)	64 (18.9)	170 (21.0)	28 (19.2)	
7.0-7.75	582 (27.9)	119 (27.7)	102 (28.1)	86 (25.4)	236 (29.2)	39 (26.7)	
7.75-8.25	581 (27.9)	140 (32.6)	95 (26.2)	92 (27.2)	220 (27.2)	35 (24.0)	
8.5-10.0	323 (15.5)	82 (19.1)	48 (13.2)	62 (18.3)	107 (13.2)	24 (16.4)	
Body-mass Index z-score: subjective							
Overweight (>85 th percentile)	378 (21.8)	93 (26.3)	76 (27.3)	58 (21.1)	131 (18.5)	20 (17.5)	0.005
Obese (>95 th percentile)	146 (8.4)	37 (10.5)	35 (12.6)	23 (8.4)	43 (6.1)	8 (7.0)	0.009
Body-mass Index z-score: objective							
Overweight (>85 th percentile)	431 (28.4)	94 (30.9)	109 (34.2)	79 (27.8)	115 (23.9)	34 (26.4)	0.02
Obese (>95 th percentile)	188 (12.4)	50 (16.5)	42 (13.2)	33 (11.6)	45 (9.4)	18 (14.0)	0.05

1. Test of significance differences across schools (chi-square or Fisher exact tests)

Discussion

Later school start times could be a durable, transformative, and complementary strategy for dealing with the current, complex public health challenge of the obesity epidemic. This study will also provide the first-ever population-based quasi-experimental research study, with objectively measured outcomes, to examine the causal effect for or against increased sleep on weight gain prevention and thus may have important implications for other types of sleep intervention strategies. At the completion of this study we will have documented how the development of weight and weight-related behaviors differs for adolescents who attend high schools that shift to a later start versus those that stay at an early start. The product of this research is expected to be a valuable resource for guiding the development of daily school schedules nationwide and thus could have a significant and far-reaching impact on adolescent obesity prevention.

Table 5. Comparison of survey and substudy participants at Baseline.

	Substudy 1	Non-substudy 1	<i>P_{diff}</i>
N	285	2137	
Height (cm, self-report)	170.0 (9.48)	169.5 (9.51)	0.4619
Weight (kg, self-report)	64.3 (14.8)	62.6 (15.3)	0.1050
BMI Z-score (self-report)	0.376 (0.929)	0.257 (1.058)	0.0858
School night sleep hours (self-report)	7.44 (1.11)	7.31 (1.16)	0.0717
Gender (%)			0.0972
Male	45.1	51.1	
Female	53.0	47.9	
Prefer not to respond	1.9	0.97	
Eligible for free/ reduced-price meals (%)			0.0081
No	73.9	67.9	
Yes	6.72	13.7	
Don't know	19.4	18.4	
1+ parent completed college (%)			0.0468
No	19.6	15.7	
Yes	74.9	81.5	
Don't know	5.5	2.8	

Acknowledgements

All study procedures were reviewed and approved by the University of Minnesota Institutional Review Board (IRB) and the school districts. This study is supported by funding from the National Institutes of Health's (NIH) Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) (R01 HD088176). The authors would like to thank both the adolescents participating in the START study as well as the districts that have welcomed us to do research in their schools. Also thanks to Bill Baker for his work to manage they data and Kath Lenk for developing tables for this manuscript. As always, many thanks to Kate Bauer for her good ideas.

References

- Alanna J, M., Rhodes, D. G., Baer, D. J., Murayi, T., Clemens, J. C., Rumpler, W. V., Paul, D. R., Sebastian, R. S., Kuczynski, K. J., Ingwersen, L. A., Staples, R. C., and Cleveland, L. E. (2008). The US department of agriculture automated multiple-pass method reduces bias in the collection of energy intakes. *The American Journal of Clinical Nutrition*, 88(2):324–332. Available from: <https://doi.org/10.1093>.
- Allison, P. D. (2012). Handling missing data by maximum likelihood. *SAS Global Forum*, 21.
- Ancoli-Israel, Roger, C., Alessi, C., Chambers, M., Moorcroft, W., and Pollak, C. P. (2003). The role of actigraphy in the study of sleep and circadian rhythms. *Sleep*, 26(3):342–392. Available from: <https://doi.org/10.1093>.
- Anon.n.d. (2015a). . “demographics.” Available from: <https://bit.ly/377CrfW>.
- Anon.n.d. (2015b). “minnesota schools data.” Available from: <https://bit.ly/2G29mqC>.
- Anon.n.d. (2015c). “schools and staffing survey (sass).” Available from: <https://bit.ly/2R48N5V>.
- Anon.n.d. (2015d). “sleep for science.” Available from: <https://bit.ly/2TvJzip>.
- Berkey and Colditz, G. A. C. S. (2007). Adiposity in adolescents: Change in actual BMI works better than change in BMI z score for longitudinal studies. *Annals of Epidemiology*, 17(1):44–50. Available from: <https://doi.org/10.1016>.
- Control, A. R. C., Sciences, P., and Recall, A. A. A. S.-A. .-H. (2012). *Bethesda, MD: National Cancer Institute, US National Institutes of Health*.
- Dimick and Ryan, A. M. J. B. (2014). Methods for evaluating changes in health care policy. *JAMA*, 312(22):2401. Available from: <https://doi.org/10.1001>.
- Edward, O’Malley, and O’Malley, M. B. (2008). Book review sleep and psychiatric disorders in children and adolescents (sleep disorders. 5.) edited by anna ivanenko. 415 pp., illustrated. new york, informa healthcare, 2008. 978-1-4200-4807-0. *New England Journal of Medicine*, 359(24):2622–2623. Available from: <https://doi.org/10.1056>.
- Fatima, Doi, Y. S. A. R., and Mamun, A. A. (2015). Longitudinal impact of sleep on overweight and obesity in children and adolescents: a systematic review and bias-adjusted meta-analysis. *Obesity Reviews*, 16(2):137–149. Available from: <https://doi.org/10.1111>.
- Fitzmaurice, Laird, G. M. N. M., and Ware, J. H. (2011). *Applied Longitudinal Analysis*. John Wiley & Sons, Inc. Available from: <https://doi.org/10.1002>.
- Foundation., N. S. (2000). Adolescent sleep needs and patterns: Research report and resource guide.

- Glaser, G., B., and Strauss, A. L. (1967). The discovery of grounded theory: Strategies for. new york: Aldine. pages 1–18. Routledge.
- Group, Working, A. S., and on School Health Committee on Adolescence, C. (2014). “school start times for adolescents.” *Pediatrics peds.2014-1697*.
- Hibbs, A. M., Storfer-Isser, A., Rosen, C., Ievers-Landis, C. E., Taveras, E. M., and Redline, S. (2013). Advanced sleep phase in adolescents born preterm. *Behavioral Sleep Medicine*, 12(5):412–424. Available from: <https://doi.org/10.1080>.
- Jennifer, G., Fritschi, L., Heyworth, J., and Waters, F. (2012). Validation of self-reported sleep against actigraphy. *Journal of Epidemiology*, 22(5):462–468. Available from: <https://doi.org/10.2188>.
- Keyes, K. M., Maslowsky, J., Hamilton, A., and Schulenberg, J. (2015). The great sleep recession: Changes in sleep duration among US adolescents, 1991-2012. *PEDIATRICS*, 135(3):460–468. Available from: <https://doi.org/10.1542>.
- Marx, R., Tanner-Smith, E. E., Davison, C. M., Uffholz, L.-A., Freeman, J., Shankar, R., Newton, L., Brown, R. S., Parpia, A. S., Cozma, I., and Hendrikx, S. (2017). Later school start times for supporting the education, health, and well-being of high school students. *Cochrane Database of Systematic Reviews*. Available from: <https://doi.org/10.1002>.
- Minges, K. E. and Redeker, N. S. (2016). Delayed school start times and adolescent sleep: A systematic review of the experimental evidence. *Sleep Medicine Reviews*, 28:86–95. Available from: <https://doi.org/10.1016>.
- n.d. Position Statements, S. S. L., on Sleep, R., and Times., S. S. (2012).
- Neumark-Sztainer, D., Story, M., Hannan, P. J., and Croll, J. (2002). Overweight status and eating patterns among adolescents: Where do youths stand in comparison with theHealthy people 2010objectives? *American Journal of Public Health*, 92(5):844–851. Available from: <https://doi.org/10.2105>.
- of Medicine (US) Committee on Sleep Medicine, I. and Research. (2006). Sleep disorders and sleep deprivation: An unmet public health problem. *edited by H. R. Colten and B. M. Altevogt. Washington (DC): National Academies Press (US)*.
- Owens, J. A., Au, R., Carskadon, M., Millman, R., Wolfson, A., Braverman, P. K., Adelman, W. P., Breuner, C. C., Levine, D. A., Marcell, A. V., Murray, P. J., and O’Brien, R. F. (2014). Insufficient sleep in adolescents and young adults: An update on causes and consequences. *PEDIATRICS*, 134(3):e921–e932. Available from: <https://doi.org/10.1542>.
- Phelan, J. C., Link, B. G., and Tehranifar, P. (2010). Social conditions as fundamental causes of health inequalities: Theory, evidence, and policy implications. *Journal of Health and Social Behavior*, 51(1_suppl):S28–S40. Available from: <https://doi.org/10.1177>.

- Taie, S., and Goldring, R. (2017). Characteristics of public elementary and secondary schools in the united states: Results from the 2015–16 national teacher and principal survey first look (nces 2017-071).
- Taie, S. and Goldring, R. (2017). Characteristics of public elementary and secondary schools in the united states: Results from the 2015–16 national teacher and principal survey. *National Center for Education Statistics*.
- Wahlstrom, K., Dretzke, B., Gordon, M., Peterson, K., Edwards, K., and Gdula, J. (2014). Later school start times for supporting the education, health, and well-being of high school students: A multi-site study.
- Wheaton, Chapman, A. G. D. P., and Croft, J. B. (2016). School start times, sleep, behavioral, health, and academic outcomes: A review of the literature. *Journal of School Health*, 86(5):363–381. Available from: <https://doi.org/10.1111>.
- Wheaton, Ferro, A. G. G. A., and and, J. B. C. (2015). School start times for middle school and high school students — united states, 2011–12 school year. *MMWR. Morbidity and Mortality Weekly Report*, 64(30):809–813. Available from: <https://doi.org/10.15585>.
- Wing, C., Simon, K., and Bello-Gomez, R. A. (2018). Designing difference in difference studies: Best practices for public health policy research. *Annual Review of Public Health*, 39(1):453–469. Available from: <https://doi.org/10.1146>.