# The Effect of School-Based Health Centers on Adolescent Mental Health and Behavior

Anjali Pai

November 5, 2024

## ABSTRACT

The past decade has seen a worsening adolescent mental health crisis paired with low rates of take-up for mental health services. This paper examines whether increased access to mental health services has meaningful impacts on adolescent mental health and behavior. Specifically, I study the effect of school-based health centers — full-service clinics located in schools that offer physical, mental, and reproductive health services at low to no cost — on suspensions and dropouts, two measures that have been hypothesized to be linked to untreated mental health issues. First, using data from a statewide survey on school climate and socioemotional well-being I provide descriptive evidence that worse reported mental health and school climate are positively correlated with higher suspension rates but not necessarily with higher dropout rates. Next, I look at the effect of access to a schoolbased health center using a difference-in-differences analysis that leverages the timing of health center openings in California and a propensity-score matched control group. The opening of a new school-based health center decreases school-level suspension rates by around 1.2 percentage points (20% of the baseline suspension rate) within 3 years of the opening when compared to matched schools. A heterogeneity analysis provides suggestive evidence that these effects are driven by decreases in suspensions from "disruptive behavior", rather than weapon possession, violence, or drug use. I find no effect on dropout rates, suggesting that the decline in suspensions is unlikely to be caused by the crowd-out of delinquent behavior by an increase in dropouts. These results suggest that school-based health centers warrant further consideration as an effective means of addressing adolescent mental health.<sup>1</sup>

<sup>&</sup>lt;sup>1</sup>Thank you to the California School-Based Health Alliance (CSBHA) and the California Department of Education (CDE) for providing the data and valuable contextual information for this project. In particular, this project would not have been possible without the support of Lisa Eisenberg, Amy Ranger, Amy Blackshaw and others at the CSBHA. Additionally, this project has benefited from conversations with numerous researchers and School-Based Health experts, including: Maryjane Puffer, Marsha Ellis, and Alex Zepeda at the Los Angeles Trust; Samira Soleimanpour at UCSF; and Jonathan Isler at the CDE. Finally, I am thankful for the feedback of my advisor, Julian Betts, my dissertation committee, seminar participants at UC San Diego, and the feedback of colleagues at Abt Global.

## 1 Introduction

The past decade has seen a worsening mental health crisis amongst adolescents. The Center for Disease Control and Prevention (CDC) reports that the fraction of adolescents across the U.S. reporting "persistent feelings of sadness or hopelessness" increased from 26% in 2009 to 36.7% in 2019 (CDC, 2022). By 2021, one year into the COVID-19 pandemic, this share had increased to 44% of adolescents, marking a 20% increase in just two years (CDC, 2022). Untreated mental health issues may have a direct impact on an individual's ability to succeed in various areas of life. A large body of research in psychology suggests that for adolescents in particular, untreated mental health issues may manifest in disruptive behavior and inattention (Garland et al. (2010), McLeod et al. (2012)). Recent quasi-experimental work suggests that behavioral disorders such as ADHD (Attention Deficit Hyperactive Disorder) in childhood may have negative effects on test scores and attainment of schooling (Currie and Stabile, 2006). Moreover, 50% of mental health illnesses identified in adults are documented as having begun before the age of 14 (Kessler et al., 2005), suggesting that treating mental health amongst adolescents may have both positive impacts on adolescent outcomes and positive spillovers on outcomes in adulthood.

Despite the high rates of reported depression amongst adolescents, take-up of mental health services continues to be relatively low. The 2021 National Survey on Drug Use and Health, administered annually by the Substance Abuse and Mental Health Services Administration (SAMHSA), reports that amongst a sample of 5 million adolescents aged 12 to 17 who reported having a "major depressive episode" in the past year, only 40% reported receiving treatment for depression (SAMHSA, 2023). The gap between adolescents' reported need for mental health services and service utilization may be indicative of a combination of supply-side and demand-side barriers. On the supply-side, for example, there is evidence that the availability of mental health services has been unable to meet the documented need for mental health treatment since at least 2016, with the fraction of documented need met decreasing from 56% in 2016 to around 27 % in 2023 (BHWF, 2023).

On the demand-side, there are three key documented barriers to take-up of mental health services: physical distance, financial cost, and societal stigma (NCBH, 2019). This paper focuses on evaluating a model of healthcare provision that is theoretically well-designed to overcome all three of these demand-side barriers: school-based health centers. School-based health centers (or SBHCs as they are commonly known) are student-focused health clinics that are located directly in or near a K-12 school, provide services at low or no-cost to students, and are tightly integrated into the school with which they are associated. These three features have the potential to address distance, cost, and stigma barriers respectively, suggesting that SBHCs may be well positioned to improve adolescent take-up of mental health services. While the first SBHCs arose with a goal of filling gaps in physical health in low-income communities, they quickly expanded to offering mental health services in the early 2000s (Flaherty and Osher, 2003). As of 2017, nearly 65% of the

SBHCs nationwide reported employing behavioral health specialists in addition to primary care providers.<sup>2</sup> While school-based health centers have been a feature of many low-income schools for several decades, there is limited causal evidence on their efficacy. The growing popularity of SBHCs and the growing push from state and federal authorities to increase funding and resources to treat adolescent mental health suggest a need for causal evidence on the effectiveness of SBHCs as a tool for addressing adolescent mental health.

This paper aims to isolate the impact of school-based health centers on mental health by examining the effect of school-based health centers in California on delinquent behavior (measured by suspension rates) and dropout behavior (measured by dropout rates), two behavioral outcomes that are most likely to be directly impacted by school-based health centers through the treatment of adolescent mental health issues. I construct a novel panel dataset that links data on the openings of all SBHCs in California between 2011 - 2019 with school-level data on suspension rates, dropout rates, and student demographics. The main threat to identifying the direct effects of a school-based health center is that the decision to open one of these centers is not random. As a result of large short-run construction costs and long-run operational costs associated with opening a school-based health center, the ability to open a center often hinges on strong community partnerships and significant buy-in from schools and school-districts. This makes it difficult to directly compare outcomes in a school with a school-based health centers to outcomes in a school without one.

To address the selection issue, I employ a propensity-score matching approach to select a control group that is most likely to be similar on observable and unobservable characteristics to the set of schools that open an SBHC. Combining this matching with a staggered-event difference-indifferences model allows me to isolate the effect of access to an SBHC by comparing the trend in outcomes following the opening of a new SBHC to the trend in the years preceding the opening. I find evidence that access to an SBHC reduces the rate of suspensions for a school by around 1.2 percentage points. This is a large magnitude effect, amounting to a 20% decrease from the control group baseline rate. I also find a declining trend in the rate of repeat suspensions that (while insignificant) suggests that access to school-based health centers could actually be affecting the patterns of behavior that lead to suspension, and not just providing schools with an alternative to suspending students. Although I find no consistent effects on dropout rates, tight 95% confidence intervals rule out increases and decreases larger than 0.5 percentage points, indicating that the decrease in suspension rates is unlikely to be explained by an increase in dropout rates crowding out suspensions. These treatment effects are robust to correcting for the possibility of bias from negative weighting in staggered-event difference-in-differences. In addition to using the corrected estimator proposed by Callaway and Sant'Anna (2021), I use a decomposition method proposed by de Chaisemartin and D'Haultfœuille (2020), to show that under my preferred specification, 95% of the weights assigned to the individual 2 x 2 difference-in-difference estimates are positive and that there is unlikely to be sufficient treatment effect heterogeneity across groups or periods to

<sup>&</sup>lt;sup>2</sup>Source: These statistics come from the 2016-2017 National Census of School-Based Health Centers (Love et al., 2019).

significantly bias my primary estimates.

Exploring the mechanisms behind these treatment effects, I find that the decrease in suspension rates is strongest for "defiance suspensions", suspensions that are caused by disruptive or defiant student behavior. Comparatively, I find no change in the fraction of suspensions that results from weapon possession, violence, or drug use.<sup>3</sup> This provides suggestive evidence that the decrease in suspensions is driven by a decrease in disruptive behavior, which is a common symptom of many underlying psychological disorders in adolescents (Garland et al., 2010). Finally, I am able to link suspension and dropout rates for a subset of schools to data from the California Healthy Kids Survey (CHKS), a biannual survey on school climate, risky behavior, and mental health, administered by California school districts. Using this survey data I show that controlling for cross-year and crossschool differences, higher rates of reported depression and lower levels of "school-connectedness" are correlated with higher suspension rates but not necessarily with higher dropout rates. This further supports the theory that changes in suspension rates may be capturing improvements in adolescent mental health outcomes, although data limitations preclude me from drawing that conclusion explicitly. In addition to providing some of the first quasi-experimental evidence that school-based health centers may decrease suspension rates, the methods I employ highlight some of the challenges to identifying causal impacts of school-based health centers that are valuable to consider for future research on this topic.

This paper contributes to an existing literature on the impacts of school-based health centers, which is predominantly descriptive and non-causal. While previous studies point to positive relationships between school-based health centers and attendance, academic performance, physical health, and graduation rates, these results come from either cross-sectional comparisons between schools with and without SBHCs (Kisker and Brown (1996), Santelli et al. (1996), Paschall and Bersamin (2018)), within-school comparisons between students who utilize SBHC services and students who do not utilize these services (Kerns et al. (2011), Walker et al. (2010), McCord et al. (1993)), or single-school program evaluations (Gall et al. (2000), Warren and Fancsali (2000)). An exception is Lovenheim et al. (2016), which takes a quasi-experimental approach to studying the effect of SBHCs on teenage fertility and high-school dropout rates. Using data from a national survey of SBHCs and a staggered-event difference-in-differences approach, the paper finds that the first opening of an SBHC in a county leads to a 1.3% decrease in the teenage fertility rate, and has no identifiable effect on high school dropout rates. These results suggest potentially large effects of SBHCs on improving reproductive health. I contribute to this literature by providing the first quasi-experimental analysis of the effect of school-based health centers on delinquency, and isolating effects at the school-level rather than the county or district levels. Although I use a similar staggered-event difference-in-differences approach to Lovenheim et al., I supplement it

<sup>&</sup>lt;sup>3</sup>The California Department of Education defines six total categories of offenses that may lead to a suspension: (1)  $V_{i}$  by the transformation of the second second

<sup>(1)</sup> Violent Incident (Injury); (2) Violent Incident (No Injury); (3) Weapons Possession; (4) Illicit Drug-Related;

<sup>(5)</sup> Other; (6) Defiance-only. Categories (1)-(4) consist of federal offenses and Category 5 consists of offenses under state law that are not against federal law. Appendix Table F.1 lists the offenses in each category.

with a propensity score matching approach in order to be able to conduct this analysis at the level of schools rather than aggregating up to the level of districts. This approach is especially useful when attempting to isolate impacts of SBHCs on mental health, since it isolates the effect for the students with lowest distance-barrier to accessing the school-based health center. A major strength of the Lovenheim et al. analysis is the ability to look at granular categories of services offered by each SBHC in their sample. Due to data limitations, I am unable to acquire such information for the SBHCs in my sample; however, I compensate for this by leveraging data on suspensions that is disaggregated by the type of offense and integrating survey data on mental health and school climate to provide suggestive evidence that my results are more likely to be driven by improvements in mental health than other mechanisms. My paper also complements recent work from Komisarow and Hemelt (2022) that focuses on a school-based telemedicine program in rural North Carolina. This paper finds that access to a "school-based telemedicine center" decreases the likelihood of chronic absenteeism by 2.5 percentage points and decreases the likelihood of a student having at least one violent or weapons-related infraction by 40-47% of the baseline mean. Although the intervention studied is very different (telemedicine as opposed to in-person services), the results support the theory that more "comprehensive" school-based health services can have a large impact on mental-health linked outcomes.

Finally, my work contributes to a smaller literature on school-based approaches to mental health provision. The causal work on this topic has largely focused on the impact of added elementary school counselors on students' behavioral outcomes. Carrell and Carrell (2006) and Carrell and Hoekstra (2014) find that increasing the ratio of counselors to students in elementary schools reduces disciplinary incidents. Similarly, Reback (2010b) finds that state reforms that improve the ratio of counselors to students in elementary schools reduce teachers-reported incidents of delinquent behavior. Finally, Reback (2010a) concludes that increased funding for elementary-school counselors has a significant impact on decreasing disciplinary infractions. An important difference between elementary school counselors and the mental health professionals staffed in school-based health centers is that school counselors do not usually provide on-site therapy or formal mental health treatment, and will instead refer students to outside services if mental health needs are identified. My paper isolates the effects of a program that involves direct treatment of mental health issues on-site, rather than just preventative counseling and educational services. Additionally, by studying elementary, middle, and high schools, I provide valuable evidence on how school-based mental health services can affect students of a wider age-range. The decrease in suspensions I identify is primarily driven by middle schools and high schools in the sample, with middle schools seeing a larger decrease than high schools. This suggests that school-based health centers may be well-targeted for middleschool and high-school aged youth, who have been understudied in this literature.

The remainder of this paper proceeds as follows: Section 2 provides background on school-based health centers in California; Section 3 provides an overview of the data and sample construction process; Section 4 discusses the empirical strategy and identifying assumptions; Section 5 shows results from my primary specifications; Section 6 shows results from a set of alternate specifications;

Section 7 shows a set of heterogeneity analyses; Section 8 discusses potential policy implications; and finally, Section 9 proposes avenues for future research and concludes.

# 2 Background

The setting for this analysis is the state of California, which serves over 5.8 million K-12 students.<sup>4</sup> The reported rate of mental health issues amongst adolescents in California has been increasing in the past decade, with nearly 45% of youths aged 12-17 reporting that they have struggled with mental health issues in 2021 (Wright et al., 2021).

The California School-Based Health Alliance (CSBHA) defines a school-based health center as a "student-focused health centers or clinics that are located on or near a school campus, are organized through school, community, and health provider relationships, and provide age-appropriate, clinical health care services onsite by qualified health professionals" (CSBHA, 2022b). Under this definition, as of the 2022-2023 school year, there were 346 active School-Based Health Centers in California.<sup>5</sup> Notably, while several U.S. states have begun to provide centralized state funding for the opening and maintenance of SBHCs, California is one of a few states that doe not offer any state funding for these centers. In schools that do not have an SBHC, health services are most commonly provided by school nurses who have the ability to assess students for health problems, deliver basic health services such as immunizations and insulin, and provide health and nutrition education; however most registered nurses do not possess the ability to treat more serious health problems, provide psychological counseling or therapy, or prescribe medication (CSHCA, 2010). In contrast, most school-based health centers are staffed by a combination of nurse practitioners, physician assistants, physicians, residents, medical assistants and nurses who are able to provide physical, mental, and reproductive health services directly on a school's campus. 77% of currently active California SBHCs report offering mental health services in addition to primary care.<sup>6</sup>

A second critical feature of school-based health centers that distinguishes them from other community health alternatives is that the care is specifically adolescent-focused and offered at low or no cost to students (CSBHA, 2022b). The offer of these services at low costs is intended to enable all students to access the services regardless of socioeconomic background or ability to pay. What this means in practice, however, is that the ability of an SBHC to fund its operations comes from the integration of various sources of funding. Most SBHCs are funded through a combination of state and local grants and community partnerships. 65% of these centers are operated by Federally-Qualified Heath Centers (FQHC), which are federally-funded, local, non-profit healthcare organizations intended to serve lower-income populations. FQHCs receive favorable Medicaid reimbursement rates, allowing them to offer services to low-income individuals at low costs (CSBHA,

<sup>&</sup>lt;sup>4</sup>Source: About School-Based Health Centers, California School Based Health Alliance

<sup>&</sup>lt;sup>5</sup>Source: Fingertip Facts on Education in California, California Department of Education

<sup>&</sup>lt;sup>6</sup>Source: School-Based Health & Wellness Centers in California: A Growing Trend, California School Based Health Alliance

2023). The other 35% of centers are funded by either local education agencies (27%), local hospital or universities (3%), local public health departments (3%), or other community-based organizations (~ 2%) (CSBHA, 2022a).

The unique features of the school-based health centers make them well-equipped to increase takeup of mental health services for adolescents by addressing the three most common barriers: physical distance, financial cost, and societal stigmas around mental illness (NCBH, 2019). The "in-school" location of these services combined with the low or no cost directly address the financial and distance cost barriers. As the least tangible of the three barriers, it is difficult to assess whether the SBHCmodel directly addresses stigma; however, there are two aspects of the SBHC model that may be well-suited to decrease societal stigma around mental health services. The first is that the behavioral health services available in SBHCS are accessible by all students in a school, increasing the likelihood of utilization spillovers within networks of peers. The second is that the model of mental health service provision followed by SBHCs in California has at its foundation, universal services that are targeted at treating and screening as many students as possible. The California School-Based Health Alliance defines three tiers of mental health service provision: Universal Prevention (Tier 1), Targeted Early Intervention (Tier 2), and Intensive Intervention (Tier 3). Appendix Figure A.1 outlines the types of services included in each tier. Most SBHCs that report offering mental health services will offer Tier 1 services at baseline, and Tier 2 and Tier 3 services depending on their staffing and available funding. For example, in the Madera South School-Based Health Center in Madera County, Tier 1 services include a program that trains students to identify mental health concerns in their peers and provide "peer counseling", while more intensive services include oneon-one counseling with licensed clinical social workers and referrals to external behavioral-health practitioners for more intensive mental health needs. At the Monroe High School Wellness Center in San Fernando Valley, the types of services range from individual and family therapy to psychiatric and psychological testing services for students.<sup>7</sup>

Beyond just targeting the expected barriers to take-up of mental health services, the adolescentfocused staffing and practices in SBHCs may give them an advantage over community health clinics in actually treating adolescents with mental health issues. A 2003 retrospective cohort study from Juszczak et al. (2003) provides non-causal evidence that adolescents with access to SBHCs have higher visit rates than students who only have access to community health centers. Moreover, for those students who do not have access to an SBHC and use only community health centers, 97% of visits were for medical services. Comparatively, for students who chose to use an SBHC, at least 34% of visits were for mental health services. While there are large concerns about selection bias in this study, the descriptive statistics it is able to provide suggest that SBHCs may provide unique access to mental health services that are either not available through community health centers or not well-targeted enough to treat adolescents. While I am not able to directly measure take-up of services in my data, discussions with administrators at school-based health centers in California

<sup>&</sup>lt;sup>7</sup>These case studies come from a report compiled by Lisa Eisenberg, formerly of the California School Based Health Alliance.

have suggested that demand for these services is high and often unmet due to the staffing and funding constraints of the centers.

# 3 Data

#### 3.1 Data Sources

This project draws on three primary data sources: (1) data on school-based health centers from the California School-Based Health Alliance, a non-profit organization that provides support and resources for school-based health centers operating in California; (2) annual data on suspensions, dropout rates, and student demographics at the school level from the California Department of Education (CDE); and (3) annual data on student-reported mental health, wellness, and behaviors from the California Healthy Kids Survey (CHKS).<sup>8</sup> The data on SBHC openings contain information on the opening dates for all "active" SBHCs in California (i.e. SBHCs that were operational as of August 2022 when the data were compiled). This sample consists of 286 SBHCs across 34 counties, 206 zip codes and 120 school districts. For each SBHC, the data contain information on: the opening date, name of the clinic, name of the associated school, address of the clinic (street address, city, county, and zip code), and a set of SBHC-reported characteristics (SBHC type, sponsoring organization, list of services provided, list of schools served, and populations served).<sup>9</sup>

The data on suspensions, expulsions and dropout rates come from the CDE's public database. The data on suspensions and expulsions include annual counts and rates for every public school in California from the 2011-12 through 2020-21 academic years, both at the school-level and disaggregated by race and gender. The data on dropout rates include similarly disaggregated annual counts and rates, but restricted to high schools in California (and in special cases for grades 7-12) from the 2010-11 to 2016-17 academic years. Finally, in an effort to connect the analysis in this paper more-directly to mental health, I incorporate data from the California Healthy Kids Survey (CHKS), a modular, anonymous assessment with well-validated pyschometric properties (Hanson and Kim (2007), Mahecha and Hanson (2020)).<sup>10</sup> The CHKS is administered in elementary, middle and high schools by school districts across California at an annual or bi-annual frequency.

Figure 1 shows the distribution of SBHC openings in California across time. The primary takeaway from this figure is that there is large variation in the number of SBHCs that open in each year. A majority of openings fall between 1990-2020, which is coincidental with the start of the nationwide "boom" in SBHCs in 1990 (Flaherty and Osher, 2003). Due to limitations in the availability of data on suspensions and dropout rates from the CDE, the primary analyses in this

<sup>&</sup>lt;sup>8</sup>The California School-Based Health Alliance is a state affiliate of the National School Based Health Alliance.

<sup>&</sup>lt;sup>9</sup>The variable reporting "schools served" is generated from an open-ended text response, which is relatively sparse in the data and therefore a sparse measure of actual SBHC service area.

<sup>&</sup>lt;sup>10</sup>These data were acquired through partnership with the California Department of Education. Many thanks to Jonathan Isler at the CDE for his assistance with acquiring this data.



Figure 1: This graph shows the distribution of SBHC openings between 1967-2023 for the set of SBHCs that were active in California as of 2022. The x-axis shows the range of opening years while the y-axis shows the total number of SBHCs that opened in that year. The yellow shaded bars mark the set of years covered by the California Department of Education's suspension and dropout data.

paper limit to the set of openings between 2012-2019, which are indicated by the yellow-shaded bars in Figure 1. The available survey data from the CHKS extends back to 1998, however when used in combination with suspension and dropouts data, I am once again forced to limit my sample to this time range.

## 3.2 Constructing the Study Sample

Constructing the analysis dataset involved a few key steps that I will discuss in this section. First, to merge the data on SBHC openings to the CDE data on school-level suspensions and dropout rates, it is necessary to first match each SBHC to a "principal school" in the CDE's "Public Schools and Districts" data. For on-site SBHCs, the "principal school" can be defined as the school in which the SBHC is physically located. For off-site and mobile vans, I define the "principal school" as the school in closest geographic proximity to the SBHC. In practice, SBHCs are matched to schools using an iterative process of fuzzy string matching on address and school name. In the SBHC data, the "school name" comes from a survey question where SBHCs were asked to report the name of the "school served". The address components come from text fields for the SBHC's reported street address, zip code, city, and county. In the CDE data, school name and complete school address are standardized fields available for all California public schools.<sup>11</sup>

The matching process begins by identifying all potential matching schools for each SBHC using the composite addresses of the SBHC and the school, and assigning each match a "similarity score" between 0 and 1, with a higher value indicating a "better match".<sup>12</sup> For each potential match, I also generate similarity scores based on the school name, gradespan, city, county, and zip code fields.<sup>13</sup> Table A.2 outlines the nine-step fuzzy string matching process that is used to select a "best match" for each school based on these calculated similarity scores. <sup>14</sup> Each subsequent stage of this iterative matching process is less stringent than the prior one, with the goal of generating matches that are as exact as possible.

At the end of this matching process each SBHC is matched to a "County-District-School" (CDS) code, which is a unique school-identifier utilized across all CDE datasets. This CDS code allows

<sup>&</sup>lt;sup>11</sup>It is worth noting that the SBHC data does not contain a field for school district. Without this information, the "school name" field on its own is insufficient for matching, since there are several instances of schools with the same name that exist in different districts (eg. "Jefferson High School" and "Lincoln High School" each show up multiple times. Moreover, since this data is collected directly from SBHCs, there is no standard convention for formatting the name of a school; therefore the same school is likely to be identified by a different name in the SBHC data than in the standardized CDE data.

<sup>&</sup>lt;sup>12</sup>Mechanically, similarity scores are generated with the *matchit* package in Stata which decomposes the text into bigrams before calculating a Jaccard similarity index.

<sup>&</sup>lt;sup>13</sup>The "gradespan" field is a standardized field in the CDE data, but does not exist in the SBHC openings data. To generate a similarity score for this field, I generate a corresponding "gradespan" field in the SBHC data by parsing the "school name" field for keywords such as "elementary", "middle", and "high".

<sup>&</sup>lt;sup>14</sup>The matching procedure is as follows: if a unique match can not be selected in the first iteration, the algorithm will proceed to the second tier of matching, and so on. This continues until a single, unique match is ident-fied for each SBHC.

each SBHC to be attached to a panel of outcomes data on suspension rates and dropout rates for its associated primary school. It also allows school-level characteristics such as racial composition, enrollment, and fraction of students on Free-or-Reduced-Price Lunch to be merged onto the SBHC dataset. These same CDS codes allow the data on suspensions and dropouts to be linked to the California Healthy Kids Survey data.

Finally, in order to eventually run an event-study analysis, the opening year of each SBHC must be identified. The data on SBHCs contains the calendar opening date for each SBHC. The outcomes from the CDE, however, are defined in terms of *academic years* each of which spans half of two consecutive calendar years. This leaves the researcher to make a choice regarding how these calendar opening dates should be assigned to academic years. The most conservative rule would assign each SBHC to the first full academic year for which it is open. This would mean that an SBHC opening anytime in calendar year 2011 would be assigned an academic opening year of 2011-2012. The most lenient possible rule would assign an SBHC to academic year y as long as the SBHC is open for one or more days of academic year y.

All main specifications in this paper follow a rule that errs on the side of leniency, and assigns an SBHC to academic year y as long as it is open for at least one full month of year y. Under this rule, an SBHC that opened in April of 2011 would be assigned an academic opening year of 2010-2011, but an SBHC that opened in June would be assigned the academic year 2011-2012. Since this rule treats schools that opened one month before the end of the school year the same as schools that opened at the start of the school year, in the main event study specifications, I assign a weight to the opening year that accounts for the fraction of the year for which the SBHC was open.

#### 3.3 Sample Restrictions

The sample of school-based health center openings is restricted in two ways. First, I only use openings between 2012 - 2018, inclusive. The lower bound on opening years is necessary since the earliest year of data on suspension and dropout rates is the 2011-12 school year.<sup>15</sup> The upper bound is imposed to avoid including schools that opened a center during or after the COVID-19 pandemic. In particular, this restriction addresses the concern that the choice to open an SBHC after 2019 may be a *direct response* to increasing mental health concerns during the pandemic, and therefore may be correlated with the imposition of other policies that target adolescent mental health.

The second restriction limits the SBHCs in the analytical sample to those that are defined as "on-site" in the data. The CSBHA data on SBHC openings contains three types of centers: on-site, off-site (which includes telehealth-only centers), and mobile vans. In the universe of SBHCopenings in California, 73% are on-site, 13% are off-site or "telehealth-only", and 14% are mobile vans. Theoretically, SBHCs located directly within a school building may be the most well-equipped

 $<sup>^{15}</sup>$ SBHCs opening in 2011 cannot be included in the sample since they would have zero years of outcomes prior to the opening.

to directly address the primary barriers to take-up of mental health services. In particular, these services would be accessible to students during the school day and without leaving their school building, which may decrease distance barriers. If school-based health centers improve take-up of mental health services in part by reducing *distance* to access services, then those effects should be strongest and most prominent for students in the school that hosts the SBHC. Furthermore, the location of a center within a school may lead to closer integration between the health center and the broader school community, which has the potential to decrease the stigma around utilizing these services.

A further reason to restrict to on-site SBHCs is that centers in this sub-sample are more likely to offer mental health services, are more homogeneous in their characteristics, and have a more easily defined "treated school" and are more likely to offer mental health services. A comparison of on- and off-site SBHCs presented in Table 1 reveals that 83% of on-site SBHCs offer mental health services, compared only 48% of non-on-site SBHCs. Additionally, on-site SBHCs are more homogenous in their community partnerships — over 80% of on-site SBHCs are sponsored either by a Community Health Center (CHC) or a school system while off-site SBHCs are sponsored by a more diverse set of organizations — which suggests that the forces motivating the opening of on-site SBHCs may be more similar than those motivating off-site SBHCs. Finally, a benefit of using on-site SBHCs is that they offer a natural definition for the primary population of students who are exposed to the treatment figure: students attending the school in which the SBHC is located. Figure 2 compares the reported service populations for the three types of SBHCs. On-site SBHCs (represented by the maroon bars) are more likely to report only serving their "principal school" and less likely to report serving the entire district or multiple districts. This observations suggests that there is lower risk to restricting the catchment area for treatment with an on-site SBHC than there may be with an off-site SBHC.

Finally, all schools and SBHCs located in Los Angeles Unified School District (LAUSD) are omitted from the analytic sample due to a district-wide ban on "willful defiance" suspensions that was enacted in 2013 (Rott, 2013). This change in suspension policy that is coincidental with the window for this study makes it difficult to disentangle changes in suspension rates over time in LAUSD due to the suspension policy from changes over time due to the opening of SBHCs. LAUSD is also the singular school district in California to enact such a policy in this time period, which would raise concerns that the SBHCs that opened in LAUSD after 2013 were driven by different motivations and goals than SBHCs opening in other districts. Robustness checks presented in Appendix E show that the primary results for overall suspension rates do not meaningfully change when LAUSD is included in the sample. Notably when LAUSD is included in the sample, the effect of SBHCs on the defiance suspensions is larger and more statistically significant than in my primary specifications, which is to be expected given the nature of the policy change.

	All SBHCs	On-Site	Off-Site/Mobile
Opening Date			
Opened Between 2011-2019	$0.38\ [0.49]$	$0.37 \ [0.48]$	$0.41 \ [0.49]$
Gradespan of Principal Linked School			
High School	$0.48 \ [0.50]$	$0.55 \ [0.50]$	$0.30 \ [0.46]$
Middle School	$0.14 \ [0.34]$	$0.15 \ [0.36]$	0.09  [0.29]
Elementary School	$0.25 \ [0.43]$	$0.23 \ [0.42]$	$0.30 \ [0.46]$
Other/Unidentified	$0.13 \ [0.34]$	$0.07 \; [0.25]$	$0.30 \ [0.46]$
Categories of Services Offered			
Mental Health	$0.73 \ [0.44]$	$0.83 \ [0.38]$	$0.48 \ [0.50]$
Medical	$0.82 \ [0.39]$	$0.83 \ [0.38]$	$0.80 \ [0.40]$
Reproductive Health	$0.62 \ [0.49]$	$0.63 \ [0.48]$	$0.61 \ [0.49]$
Dental or Vision	$0.62 \ [0.49]$	$0.60 \ [0.49]$	$0.66 \ [0.48]$
Categories of Populations Served			
Serves Other Students	$0.42 \ [0.49]$	$0.38\ [0.49]$	$0.53 \ [0.50]$
Serves Other Youth	$0.56\ [0.50]$	$0.51 \ [0.50]$	$0.68 \ [0.47]$
Serves Community	$0.44 \ [0.50]$	$0.38\ [0.49]$	$0.62 \ [0.49]$
Serves Families	$0.59\ [0.49]$	$0.57 \ [0.50]$	$0.65 \ [0.48]$
Primary Sponsoring Organization			
CHC Sponsored	$0.52 \ [0.50]$	$0.52 \ [0.50]$	$0.52 \ [0.50]$
Hospital Sponsored	$0.04 \ [0.20]$	$0.02 \ [0.14]$	$0.10 \ [0.30]$
Health Department Sponsored	0.08  [0.27]	$0.05 \ [0.22]$	$0.14 \ [0.35]$
School System Sponsored	$0.26 \ [0.44]$	$0.31 \ [0.47]$	$0.13 \ [0.33]$
Private Nonprofit Sponsored	$0.08 \ [0.27]$	$0.07 \ [0.26]$	$0.10 \ [0.30]$
Other Sponsored	$0.02 \ [0.14]$	$0.02 \ [0.15]$	$0.01 \ [0.11]$
Observations	286	207	79

 Table 1: Summary Statistics: All SBHCs, On-Site, and Off-Site

Standard deviations in brackets



Figure 2: The figure graphs the fraction of school-based health centers within each center "type", that report serving each of the following areas: the principal school (i.e. the school to which the SBHC is attached), multiple schools, the entire district, or multiple districts. Data on the schools served comes from a self-reported text field where the school-based health center provides a list of all schools it serves. I determine the four categories as follows: SBHCs that only list one school, where the school matches the principal school are classified as serving the "principal school"; SBHCs that list the name of more than one school are classified as serving "multiple schools"; SBHCs that list their principal district"; and SBHCs that list multiple districts are classified as serving "multiple districts".

#### 3.4 Outcomes: California Healthy Kids Survey

The California Healthy Kids Survey (CHKS) is a part of the California School Climate, Health, and Learning Surveys (CalSCHLS) system, which was designed to provide schools with "quality local data which can be used to improve student academic performance and social-emotional, behavioral, and physical health of all youth".<sup>16</sup> For the purposes of this study, I acquired annual CHKS datasets from the 1998-1999 school year through the 2021-2022 school year. The data is anonymous and at the student-level, and contains responses to every question from the core CHKS module, as well as from any supplementary modules the student completed. The data also provides information on the student's demographic characteristics, school, district, and grade. Districts that administer the CHKS are *required* to administer the survey in 7th and 9th grades, but are encouraged to administer it to 5th and 11th grade students as well. For the purposes of this study, I restrict to sampled 7th and 9th grade students to avoid the inclusion of grades that are not consistently sampled across all districts and schools.

My primary use of the CHKS data in this current paper is to construct measures of school climate and mental health. For each survey question used to construct one of these measures, the individual student responses are averaged to generate a "mean" school-level response. To measure school climate, I construct the following four psychometrically validated indices proposed by researchers at WestEd, the agency that developed the California Healthy Kids Survey: (1) Caring Staff-Student Relationships; (2) School Connectedness; (3) Delinquency; (4) Substance Use at School.<sup>17</sup>

Indices 1 and 2 are based on a set of questions related to positive relationships between students and school staff and feelings of belonging in a school. Examples of sentiments captured by Index 1 are "At school, there is a teacher or adult who really cares about me" and "There is a teacher or adult who believes that I will be a successful student". Examples of the sentiment captured by Index 2 are "I feel close to people at this school" and "I am happy to be at this school". These indices are selected as the ones most likely to capture sentiments that may be linked to *positive mental health*. Indices 3 and 4 are based on sets of questions surround an individual student's own delinquent behavior and substance use at school. For example, questions under Index 3 may ask the student about the frequency at which they have been "in a physical fight at school" or "carried a gun at school" or "been threatened with harm or injury at school". Questions under Index 4 focus on asking students about their own use of illicit substances at school, including cigarettes, smokeless tobacco, alcohol, and marijuana. These indices are selected as the ones most likely to to capture aggressive or destructive behavior, which may be linked to *negative mental health*. Appendix Section F.2 describes the data cleaning and index construction process for the CHKS outcomes, as well as the specific questions included in each index.

<sup>&</sup>lt;sup>16</sup>https://calschls.org/about/the-surveys/

<sup>&</sup>lt;sup>17</sup>Each index is a weighted average of the responses to a set of questions. I use the exact questions and weights suggested in Mahecha and Hanson (2020), the paper that proposes these measures and validates their psychometric properties.

For measure of mental health status, I use two questions from the CHKS that are based on commonly-used survey questions that target mental health:<sup>18</sup>

- During the past 12 months, did you ever seriously consider attempting suicide? (Yes/No)
- During the past 12 months, did you ever feel so sad or hopeless almost every day for two weeks or more that you stopped doing some usual activities? (Yes/No)

Since these are yes or no questions, I construct measures for the fraction of students in each school who responded "yes" to each question. These measures can be loosely viewed as a proxy for the fraction of students in a school who are experiencing some kind of mental health issue. The data for Question 1 is available going back to 2010, while the data on Question 2 is only available after 2014, which is when that question was added to the core module.

#### 3.5 Outcomes: Suspensions and Expulsions

The data on suspensions and dropout rates comes from the California Department of Education's public data repository.<sup>19</sup> The CDE provides annual school-level data on suspension rates from the 2011-12 through 2020-21 academic years. To construct a dataset that is compatible with the CDE-code assigned SBHC data, I aggregate the datasets for years 2011-12 through 2018-19.<sup>20</sup> In addition to overall suspension rates, the CDE data provides suspension rates disaggregated by gender and race, as well as suspension counts for 6 categories of offenses. The final dataset contains measures of: the school-level suspension rate, suspension rate for female students, and suspension rate for male students. Additionally, I use the provided counts and a variable for cumulative enrollment to construct measures of suspension rates for all six categories of offense type. Spatially, the data on suspension cover 1,031 school districts and 11,040 schools in California.

The data on dropout rates is similarly structured to the data on suspensions, but available for a shorter timespan. The CDE offers school-level data on dropout counts and total enrollment for grades 9 through 12, from the 2010-11 through the 2016-17 school years. For certain schools that serve a larger span of grades in addition to grades 9-12, dropout rates are also available for grades 7 and 8. Again, to construct a dataset compatible for merging with the data on SBHC openings, I append the individual datasets for years 2010-11 through 2016-17 and construct the following outcome measures: high school dropout rate (i.e. the total number of dropouts for grades 9-12, divided by the total number of enrolled students in grades 9-12); middle school dropout rate (i.e. the total number of dropouts for grades 7 and 8, divided by the total number of enrolled students in

(https://www.cde.ca.gov/ds/ad/downloadabledata.asp). The data on suspensions can be found under the sub-page for *Discipline* and the data on dropout rates can be found under the sub-page for *Graduate and Dropout*.

 $<sup>^{18}</sup>$  These specific questions draw from other surveys such as the Youth Behavioral Risk Factor Survey administered by the Center for Disease Control and Prevention.

<sup>&</sup>lt;sup>19</sup>This data can be downloaded from the CDE website's downloadable data files page

 $<sup>^{20}\</sup>mathrm{Observations}$  for years after 2019 are dropped due to the confounding COVID-19 pandemic.

grades 7 and 8); high school and middle school dropout rates for female students; and high school and middle school dropout rates for male students.<sup>21</sup> For all final analyses, I define a combined dropout rate, which imputes the "middle school" dropout rate for those schools with a missing dropout rate for grades 9-12. This increases power by including the number of schools included in the analysis. I confirm that limiting to dropout rates defined for grades 9-12 leads to similar magnitude results with larger confidence intervals due to the lower sample size.

Appendix table A.1 compares the means of a set of suspension and dropouts outcomes between schools that ever have an SBHC ("Treated") in the years before their SBHC opens and schools that never have an SBHC ("Untreated") across all years of data. This comparison reveals that suspension rates are significantly different between schools that ever have an SBHC and schools that never have an SBHC, which suggests that the pool of all schools without an SBHC may not be an appropriate control for the pool of schools that open an SBHC. This observations motivates a more careful choice of control group for this analysis.

# 4 Empirical Design

To identify the effect of opening a school-based health center on student outcomes, this paper leverages the "staggered" timing of center openings for a staggered-treatment difference-in-differences model. The simplest version of this design would compare schools that open an SBHC to schools that do not open an SBHC in the years before and after the opening; however, the validity of this design relies on the assumption that behavioral outcomes in treated and control schools would have evolved in parallel in absence of the school-based health center. In the current context, we might expect this assumption to be violated if there are school or district characteristics that are correlated with both the decision to open a school-based health center and student behavior. Since the decision to open a school-based health center is non-random and often driven by school, community, and district partnerships, there is reason to believe that schools that open a center may be meaningfully different from schools that do not open a center.

To address the non-randomness in the decision to open a center, I select control schools using a propensity score matching procedure that matches schools that open an SBHC in a given year, y to schools that never open a center, but have a similar predicted likelihood of opening an SBHC in year y. The propensity score is based on two school-level factors with theoretically strong predictive power for the opening of an SBHC: the fraction of low socioeconomic status students and school size. Conversations with SBHC practitioners suggest that two of the primary goals of SBHCs are to serve as many students as possible and fill the gap in healthcare provision in communities that lack sufficient medical services. As a result, we might expect centers to open in lower-income schools and

<sup>&</sup>lt;sup>21</sup>Note that for gender or race category  $\mathbf{Y}$  and gradespan  $\mathbf{X}$ , the dropout rate is defined as the *total number of* dropouts across gradespan X with demographic Y, divided by the *total number of enrolled students* in gradespan X with demographic Y.

in larger schools with more space to accommodate an on-site clinic and the potential to improve access for a larger number of students. As is common in education research, I proxy for the fraction of low socioeconomic-status students using the fraction of students eligible for Free-and-Reduced-Price meals and proxy for school size with total school enrollment.

In addition to matching schools based on their propensity scores in the year of the relevant SBHC-opening, I restrict the sample of potential control schools from which a match is selected in two key ways. First, I limit the sample to schools that come from a district that is "open" to having SBHC, as measured by the opening of at least one SBHC in the district within 5 years of the first opening in the sample and prior to the year of matching. Second, for each school I restrict the matching sample to schools within the same gradespan (elementary school, middle school, and high school) to address the expectation that the measurement of outcomes, implementation of school-based health centers, and potential alternate policies implemented in place of SBHCs, are likely to differ across gradespans.

The former restriction addresses a small body of literature that has suggested that the sensitivity of propensity-score matching models to the choice of predictors, model specification, and the choice of control groups (Smith and Todd, 2005) can be addressed by matching within the same "local labor market" and using consistently measured dependent variables in the treated and control schools (Heckman et al., 1997) The optimal "local labor market" in this setting would be a schooldistrict or Local Education Agency (LEA); however there are two concerns with that approach. First, finding a good within-district match is not always possible outside of large school districts (eg. Los Angeles Unified); therefore, an approach that is limited to only within-district matches is likely to favor large school-districts, posing a threat to the the external validity of any identified effects. The second issue is that if the process of selecting a school to house an SBHC is non-random, schools that are un-treated in the same district and gradespan may not be appropriate matches for schools that are selected to house an SBHC. For example, if the decision of which school in a district receives an SBHC is motivated by school-specific trends in unobservable variables, the parallel trend assumption may not hold for pairs of schools matched within the same district. In fact, Appendix C shows that using a control group selected through propensity score matching within district does not satisfy the parallel pre-trends test. Given these constraints, I relax the requirement of matching within school district and instead expand the pool of schools to schools in districts with a similar "openness" to the SBHC-model.<sup>22</sup> The limitation to only districts that have opened an SBHC in the past maintains some of the benefits of matching within a local labor-market by choosing schools from districts that are likely to have similar attitudes toward school-based health interventions and are therefore less likely to have implemented alternate policies that would contaminate our estimates of the control students' outcomes.

 $<sup>^{22}</sup>$ Smith and Todd (2005) notes that while geographic restrictions are important in matching, they can be reasonably relaxed when the propensity score matching is combined with a difference-in-differences estimation, as it is here. This insight suggests that combining propensity-score matching with difference-in-difference models allows for fixed differences in outcomes between treated units and matched control units, as these would be differenced out. This approach also weakens the need to match locally.

The final matching procedure generates a sample of 38 SBHCs with 2-3 matched control schools each (yielding a total analysis sample of 148 schools). Table 2 shows the balance of the three baseline characteristics that are most likely to be related to SBHC openings, measured one period before an SBHC opening for the final matched sample. Estimates are the differences between sample means (control mean - treated mean) with p-values from a t-test that the difference is significantly different from 0 in parentheses. The control group mean is noted in brackets next to each difference. Column (1) shows these differences for all matched pairs, while Columns (2), (3), and (4) show the differences for the sub-samples of elementary, middle, and high school pairs respectively. Notably, across all columns there are no statistically significant differences in the fraction of FRPM students and total enrollment (the two variables used to construct propensity scores). Looking to variables that may be correlated with the decision to open an SBHC but are not use in matching, there is no statistically significant imbalance in the fraction of URM students in the full matched sample; however, there is a small imbalance within the matched middle school sample. The magnitude of this imbalance is around 15% of the control mean, which is small but not negligible. To control for any residual imbalances, my preferred specification controls for these school-level characteristics.

	All Matched	Elementary	Middle	High
	Diff [Contr $\mu$ ]	Diff [Contr $\mu$	Diff [Contr $\mu$ ]	Diff [Contr $\mu$ ]
Fraction of Students on FRPM	0.024 [0.727]	-0.003 [0.896]	$0.050 \ [0.656]$	-0.012 [0.654]
	(0.572)	(0.934)	(0.443)	(0.872)
Total Enrollment	-84.161 [1015.655]	$17.879 \ [596.333]$	-135.868 [1373.132]	-42.083 [802.792]
	(0.558)	(0.799)	(0.584)	(0.742)
Fraction Minority	$0.04 \ [0.80]$	-0.04 [0.89]	$0.12^{*}$ [0.78]	-0.03 [0.72]
	(0.249)	(0.240)	(0.038)	(0.708)
Number of schools	148	44	72	32

 Table 2: Difference in Sample Means (Control - Treatment)

Coefficients represent the difference between the control schools' sample mean and the treatment schools' sample mean. Control group mean is printed in brackets. *p*-values in parentheses.

Table 3 shows summary statistics of the suspension and dropout outcomes examined in this paper, in the year prior to the opening of the relevant SBHC. There are no statistically significant differences in outcomes in the period before an SBHC opens. This provides further evidence that the propensity score matching process has yielded comparable treated and control groups.

For all analyses in this paper, I present both difference-in-differences and event-study estimates implemented using the adjustment for bias in staggered difference-in-differences models proposed by Callaway and Sant'Anna (2021). The two-way fixed-effects specification that these event studies are based on is presented in Equation 1

	Control	Treated	p-value
Suspension Rate	0.06	0.07	0.098
	[110]	[38]	
Female Suspension Rate	0.03	0.05	0.058
	[110]	[38]	
Male Suspension Rate	0.08	0.10	0.134
	[110]	[38]	
Defiance Suspension Rate	0.02	0.03	0.092
	[110]	[38]	
Non-Defiance Suspension Rate	0.04	0.05	0.318
	[110]	[38]	
Violence Suspension Rate	0.05	0.07	0.408
	[110]	[38]	
Weapons Suspension	0.00	0.00	0.983
	[110]	[38]	
Drug-Posession Suspension Rate	0.01	0.02	0.028
	[110]	[38]	
Dropout Rate	0.01	0.01	0.515
	[57]	[22]	
Female Dropout Rate	0.01	0.01	0.383
	[57]	[22]	
Male Dropout Rate	0.01	0.01	0.650
	[57]	[22]	

**Table 3:** Treatment versus Control Sample Means of Outcome Variables One Period Before an SBHC
 Opening

\_

p-values are from a t-test that the treated and un-treated school means are equal Number of schools listed in brackets under means

$$Y_{st} = \alpha + \gamma_s + \delta_0 Treated_s + \sum_{\substack{\tau = -3\\\tau \neq -1}}^3 D_t^\tau + \omega_\tau \sum_{\substack{\tau = -3\\\tau \neq -1}}^3 \delta_\tau (Treated_s \times D_t^\tau) + \mu \mathbb{X}_{st} + \varepsilon_{st}$$
(1)

where  $Y_{st}$  is the dependent variable of interest,  $Treated_s$  is a dummy equal to 1 if school s is a treated school, and  $D_t^{\tau}$  is a dummy equal to 1 if the observation is  $\tau$  years after (or before if  $\tau$  is negative) the opening year for its matched pair.  $\omega_{\tau}$  is equal to the fraction of the "opening year" for which each SBHC is open if  $\tau = 0$ , equal to 0 for  $\tau < 0$  and equal to 1 for  $\tau > 0$ . The purpose of  $\omega_{\tau}$ is to prevent misestimation of the coefficient for event year  $\tau = 0$  that may occur due to differences in SBHC opening timings within an "academic year".<sup>23</sup>  $\gamma_s$  is a set of fixed-effects for each school.  $\mathbb{X}_{st}$  is a vector of school-level characteristics that includes the fraction of FRPM students, the total enrollment, and the fraction of URM students for school s in year t. Standard errors are clustered at the school level, which follows standard difference-in-differences guidance to cluster at the level at which the policy is implemented.

For difference-in-differences estimates, I follow the standard differences-in-differences version of this specification:

$$Y_{st} = \alpha + \gamma_s + \nu_t + \beta \left( Treated_s \times Post_t \right) + \mu \mathbb{X}_{st} + \varepsilon_{st}$$
<sup>(2)</sup>

where  $Y_{st}$  and  $Treated_s$  are defined as above.  $\gamma_s$  is a set of school fixed effects and  $\nu_t$  is a set of year fixed effects (controlling for differences between groups and time periods respectively).  $Post_t$  is a dummy equal to 1 if year t is after the opening year of the SBHC (including the opening year itself).  $X_{st}$  is the same vector of school characteristics defined above. As before, standard errors are clustered at the school level.

With a propensity-score matched control group, the validity of these estimates now relies on conditional parallel trends. Specifically, conditional on having similar predicted likelihoods of opening an SBHC, we should expect the outcomes for treated schools and matched control schools to evolve similarly in the absence of treatment. There are two contextual reasons to believe that given two schools with similar predicted likelihoods of opening an SBHC, the actual opening of an SBHC in school s in year y is plausibly random. First, the standard timeline for constructing an SBHC can take around 2-3 years and may vary by district, leading to randomness in the length of time between a district or school's decision to open an SBHC and the actual opening. Secondly, anecdotal evidence from school-based health administrators suggests that the most common reason for opening these centers is *physical health* concerns in the community rather than mental health concerns. Therefore, even if we do not believe that the timing of an SBHC opening is perfectly exogenous with regard to all student-level outcomes, there is reason to believe it may be exogenous with relation to the mental health and behavioral outcomes examined in this paper.

 $<sup>^{23}</sup>$ In practice, this is implemented by weighting all observations in the year of an SBHC opening by the (Opening Month)/12. All other observations are assigned a weight of 1.

The process of selecting propensity score predictors, the functional form of the propensity score, and the technical implementation of the propensity-score matching are discussed in further detail in Appendix B. In Section 5 I show that under my preferred propensity-score matching approach, key school-level characteristics are balanced between treated and control schools and the test for parallel trends in the pre-event period is satisfied. Appendix C compares results across alternate control groups and presents additional arguments for why the control group selected in my primary analysis may provide the best counterfactual given the limitations of the data. Finally, Appendix F shows robustness for the main results with standard two-way fixed effect specifications, and Appendix E shows robustness to an expanded sample of matched schools that includes LAUSD.

## 5 Results

### 5.1 Mental Health Correlations

The connection between mental health issues and suspendable behavior or dropout decisions is one that has been understudied in the existing literature. To motivate the use of these outcomes as proxies for the mental health impact of SBHCs, it is helpful to examine the relationship between suspensions, dropouts, and student mental health. To do so, I rely on a set of school climate indices constructed from the California Healthy Kids Survey. I focus on a set of four indices: delinquency, substance use, caring staff-student relationships, and school connectedness. The first two indices capture information about student behavior that we might expect to be correlated with suspensions but less obviously correlated with dropouts. The latter two indices capture information about students' perceptions of school climate and their feelings of "belonging" within their school. The "feelings" captured by the latter indices may be correlated with feelings of sadness or hopelessness that are often associated with depression. Appendix F.2 outlines the variables included in each index.

Table 4 presents regressions where the dependant variable is a Z-score of the average school-level suspension rate, normalized within the relevant analysis sample. The independant variable in each regression is the average value of that CHKS index at the school-level. For each regressor, the first column controls only for calendar-year fixed effects and the second column adds school fixed effects. The preferred model is the one that controls for both year and school fixed-effects ("Year/School FE") since this estimates the relationship between suspension rates and school climate measures after controlling for any cross-year and cross-school differences.

Examining student-reported behaviors first, Columns (1) - (4) of Table A.3 show that suspension rates are positively correlated with higher levels of delinquency and substance use. Turning to school climate, Columns (5) - (8) reveal that suspension rates are negatively correlated with higher levels of caring staff-student relationships and school connectedness, both of which are viewed as indicators of positive school climate. This suggests that schools with higher suspension rates may have worse

	Delinquency		Substance		Caring Staff		School Connectedness	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Year FE	$\rm Year/School \; FE$	Year FE	$\rm Year/School \; FE$	Year FE	$\rm Year/School \; FE$	Year FE	$\rm Year/School \; FE$
Delinquency (1-5)	1.682***	$0.378^{***}$						
	(0.127)	(0.074)						
Substance Use (1-5)			$0.667^{***}$	$0.127^{***}$				
			(0.091)	(0.047)				
Worse Caring Staff-Student Relationships (1-5)					$1.112^{***}$	$0.161^{***}$		
					(0.047)	(0.051)		
Worse School Connectedness (1-5)							$1.100^{***}$	$0.259^{***}$
							(0.037)	(0.040)
Constant	-0.810***	$0.146^{**}$	-0.366***	0.272***	-3.011***	-0.075	-2.269***	-0.215**
	(0.099)	(0.067)	(0.114)	(0.069)	(0.148)	(0.160)	(0.093)	(0.104)
Observations	10878.000	10436.000	10881.000	10438.000	10882.000	10441.000	10882.000	10441.000
Sample Mean	0.000	0.012	-0.000	0.012	-0.000	0.012	0.000	0.012

 Table 4: Correlations Between School Climate and Z-Scored Suspension Rates

Standard errors in parentheses. Observations are at the school level. \* p<0.1, \*\* p<0.05, \*\*\* p<0.01

perceived school climate or a larger fraction of students who do not feel that they are supported in their school.

Table 5 shows the same regressions specifications, where the independent variables are now the fraction of students who report that they have considered attempting suicide in the past 12 months (Columns 1 and 2) and the fraction of students who report that they have experienced depression in the past 12 months (columns 3 and 4). These results indicate that suspension rates are positively correlated with both of these measures of poor mental health status. The magnitude of the correlations is lower in part due to a smaller sample of years for which these mental health metrics are available, and the sparseness of the data for these measures even in years where they were elicited on the survey.

	Considered Suicide		Exerienced Depression	
	(1)	(2)	(3)	(4)
	Year FE	$\rm Year/School \; FE$	Year FE	Year/School FE
Fraction of Students Considered Suicide	$0.590^{**}$	0.208		
	(0.237)	(0.363)		
Fraction of Students Experienced Depression			$0.872^{***}$	$0.287^{**}$
			(0.128)	(0.142)
Constant	$0.363^{***}$	$0.417^{***}$	-0.049	$0.097^{**}$
	(0.080)	(0.093)	(0.047)	(0.043)
Observations	4809.000	3759.000	8819.000	8348.000
Sample Mean	-0.000	0.021	0.000	0.010

Standard errors in parentheses. Observations are at the school level. \* p<0.1, \*\* p<0.05, \*\*\* p<0.01

Tables 6 and 7 show regressions of dropout rates on the same set of school climate measures and mental health measures respectively. Examining Table 6, a first observation is that the magnitude of the relationships between measures of school climate and dropout rates is much smaller than the magnitude of the relationships between school climate and suspension rates. In all specifications with only year fixed effects, the direction of the relationships match the direction of the relationships for suspension rates. The addition of school fixed effects, however, seems to shift all coefficients toward zero, and in the case of caring staff-student relationship and school connectedness, even suggests a small but positive relationship.

	Delinquency		Substance		Caring Staff		School Connectedness	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Year FE	Year/School FE	Year FE	$\rm Year/School \; FE$	Year FE	$\rm Year/School \; FE$	Year FE	$\rm Year/School \; FE$
Delinquency (1-5)	0.870***	0.151**						
	(0.097)	(0.074)						
Substance Use (1-5)			$0.354^{***}$	0.010				
			(0.064)	(0.032)				
Worse Caring Staff-Student Relationships (1-5)					$0.335^{***}$	-0.157***		
					(0.072)	(0.053)		
Worse School Connectedness (1-5)							$0.451^{***}$	-0.127**
							(0.039)	(0.059)
Constant	-0.465***	0.106	$-0.213^{**}$	$0.213^{***}$	-0.831***	$0.713^{***}$	-0.928***	$0.543^{***}$
	(0.102)	(0.102)	(0.101)	(0.080)	(0.226)	(0.188)	(0.111)	(0.188)
Observations	8169.000	7761.000	8172.000	7765.000	8173.000	7767.000	8175.000	7769.000
Sample Mean	-0.000	-0.009	0.000	-0.009	0.000	-0.008	-0.000	-0.008

#### Table 6: Correlations Between School Climate and Z-Scored Dropout Rates

Standard errors in parentheses. Observations are at the school level. \* p<0.1, \*\* p<0.05, \*\*\* p<0.01

Looking instead at mental health measures in Table 7, there is once again a similar pattern. Column (3) reveals a positive and significant relationship between depression and dropout rates; however the addition of school fixed effects makes the relationship statistically insignificant and attenuates the magnitude toward zero.

	Considered Suicide		Exerienced Depression	
	(1)	(2)	(3)	(4)
	Year FE	Year/School FE	Year FE	Year/School FE
Fraction of Students Considered Suicide	-0.149	0.157		
	(0.284)	(0.182)		
Fraction of Students Experienced Depression			$1.311^{***}$	-0.463
			(0.289)	(0.428)
Constant	$0.247^{***}$	$0.168^{***}$	-0.347***	0.193
	(0.079)	(0.050)	(0.078)	(0.123)
Observations	2865.000	2642.000	5271.000	4752.000
Sample Mean	0.000	-0.019	0.000	-0.015

#### Table 7: Correlations Between Mental Health and Z-Scored Dropout Rates

Standard errors in parentheses. Observations are at the school level. \* p<0.1, \*\* p<0.05, \*\*\* p<0.01

While these regressions only provide suggestive correlations, there are a few valuable takeaways from this analysis. The first is that there seems to be a strong and consistent positive correlation between higher suspension rates and lower values of school climate indices that capture how comfortable and inter-connected students feel in their school. This connection is important if, for example, the opening of an SBHC offers students a safe space to share their negative thoughts, feelings, and emotions. In this regard, impacts of an SBHC on suspension rates could in part be driven by an overall improvement in perceived school climate.

The second takeaway is that higher suspension rates are positively correlated with higher rates of reported depression and suicidal thoughts. This supports the theory that the behaviors that lead to suspensions may be driven in part by un-treated mental health issues. This is valuable motivation for using suspension rates as a relevant outcome for studying the impacts of SBHC-access on mental health. Finally, it is worth noting that the relationships between dropout rates and both school climate and mental health are less obvious. While this may be in part due to the smaller sample size of the dropout rates data, it does suggest that even if access to an SBHC has impacts on adolescent mental health, we may not expect to see those impacts reflected in dropout rates.

#### 5.2 Suspension Rates

To identify the effect of opening an SBHC on suspension rate, I estimate a staggered difference-indifferences regression both as an event study to show trends in impacts over time, and as a two-period difference-in-differences to show average changes in outcomes before and after the opening. To address the possibility of negative weighting posited by recent work from Goodman-Bacon (2021), Callaway and Sant'Anna (2021), and de Chaisemartin and D'Haultfœuille (2020), the primary specifications shown use the Callaway and Sant'Anna (2021) adjustment.<sup>24</sup> Appendix F shows the corresponding two-way fixed effects regressions results. The two-way fixed effects regressions consistently return results of a similar (but slightly attenuated) magnitude and larger standard errors.

Table 8 shows coefficients from three variations of the event study specification presented in Equation 1, where the dependent variable is the school-level suspension rate. Specification (1) includes no additional controls, specification (2) adds school-level controls for the fraction of FRPM students, fraction of minority students, and total enrollment, and specification (3) adds linear time trends for each gradespan<sup>25</sup> to test for the robustness of these results to gradespan-specific changes over time in the outcome. Specification (2) is the preferred specification as it controls for the school-level characteristics that are most likely to be correlated with suspension and dropout rates. All event study plots show the coefficients on the interaction between *Treatment* and the event-time dummies for Specification (2).

All columns of Table 8 show an insignificant and close to 0 effect of the treatment in the pre-event years, suggesting that the parallel pre-trends test is satisfied. Looking to the post-event coefficients, Column (2) shows a decrease of between 1.2 - 1.5 percentage points in the three years following the SBHC opening (with the largest magnitude decrease occurring 2 years after the opening). Across all three specifications, the magnitude of the post-event coefficients is stable; however the individual coefficients fail to achieve statistical significance and an F-test that the joint effect across all treatment periods is zero yields an insignificant p-value of around 0.5. The sharp, visible decrease after year 0 suggests that while there may be a true negative effect, the analysis is under-powered to identify any statistically significant differences.

 $<sup>^{24}</sup>$ In practice, these regressions are run using the *csdid* command in Stata. In Columns (5) and (6), the inclusion of school-characteristics and gradespan time-trends respectively are used for generating inverse propensity weights.

 $<sup>^{25}</sup>$ Mechanically, Specification (3) adds separate fixed effects for elementary schools, middle schools, and high schools, each interacted with a continuous variable for calendar year

	(1)	(2)	(3)
	Baseline	Demographics	Gradespan TT
Treated x ( $\tau = -3$ )	-0.0017	-0.0001	-0.0056
	(0.0080)	(0.0082)	(0.0097)
Treated x ( $\tau = -2$ )	-0.0063	-0.0066	-0.0062
	(0.0060)	(0.0061)	(0.0076)
Treated x ( $\tau = -1$ )	ref.	ref.	ref.
Treated x ( $\tau = 0$ )	-0.0085	-0.0091*	-0.0119*
	(0.0053)	(0.0054)	(0.0071)
Treated x ( $\tau = 1$ )	-0.012	-0.012	-0.018*
	(0.008)	(0.008)	(0.010)
Treated x ( $\tau = 2$ )	-0.014	-0.015	-0.019
	(0.010)	(0.011)	(0.014)
Treated x ( $\tau = 3$ )	-0.015	-0.016	-0.015
	(0.010)	(0.011)	(0.015)
F-Stat/Chi-Stat	3.090	3.161	4.123
p-value	0.543	0.531	0.390
Pre-Period Control Mean	0.065	0.065	0.067
Observations	867	867	812

 Table 8: Event Study: Suspensions Rates

Standard errors in parentheses. Observations are at the school level.

F-stat and p-value come from a test that the coefficients on

Treatment X Event-Time for all post-event years are jointly equal to 0.

\* p<0.1, \*\* p<0.05, \*\*\* p<0.01

Figure 3 shows a standard event-study figure that plots the coefficients on the interaction between the dummy for a treated school and the dummies for each event-time year. While the post-event coefficients are statistically insignificant, there is a visible declining trend in suspension rates beginning one year after the opening, suggesting a true effect that is underpowered in the event study specification. Appendix Figure A.3 shows separate event study graphs for treated and control schools using the relevant two-way fixed effect specification to confirm that the decrease in the post-event period is driven by a decrease in suspensions rates in treated schools, relative to a stable suspension rate in control schools.



Figure 3: This figure plots the  $Treatment \times Event Time$  coefficients from an augmented event study where the outcome is the average suspension rate. The regression depicted controls for school fixed effects and a vector of school characteristics that includes fraction of Free and Reduced Price Meal (FRPM) students, fraction of underrepresented minority students, and total school enrollment. All lags prior to event time -3 and all leads after event time 3 are dropped from the estimation sample. Standard errors are clustered at the school level.

In order to estimate the average effect of an SBHC opening in the post-opening period, I run difference-in-differences versions of the three event study regressions from Table 8. All columns show the group averages calculating using the Callaway and Sant'Anna (2021) adjustment for staggered difference-in-differences.<sup>26</sup> The preferred specification in column (2) shows that for treated schools, suspension rates decrease by an average of 1.3 percentage points following an SBHC opening. This

<sup>&</sup>lt;sup>26</sup>Specifically, I show the *group average* constructed from the Callaway & Sant'Anna disaggregated 2 x 2 DiD estimates. Mechanically, the group average is calculated by first calculating the average effect of treatment on the treated (ATT) for each "cohort" of treated schools, and then taking the average of those group ATTs. This es-

treatment effect is significant at the 5% level. As before, the magnitude of the effect is stable across all three specifications. Given a pre-event control group mean suspension rate of around 6.5%, this amounts to a 20% decrease, from the expected baseline rate.

	(1)	(2)	(3)
	Baseline	Demographics	Gradespan $\mathrm{TT}$
Treated X Post	-0.0130**	-0.0136**	-0.0171**
	(0.0057)	(0.0059)	(0.0074)
Pre-Period Control Mean	0.0649	0.0649	0.0668
Observations	867.000	867.000	812.000

Table 9: Suspension Rates: Difference-in-Differences

Standard errors in parentheses. Observations are at the school level.

\* p<0.1, \*\* p<0.05, \*\*\* p<0.01

These results suggest that the opening of a school-based health center leads to a decline in the fraction of students suspended that persists for at least three years.<sup>27</sup> Appendix Figure A.9 reveals that the results are identical (although noisier) when restricting to only the sample of SBHCs that explicitly report offering mental health services.<sup>28</sup> While intuition suggests that mental health treatment is the most likely channel through which SBHCs might impact suspensions, it is possible that treatment of certain physical health issues may affect certain behaviors, (such as substance abuse), that could lead to suspension. Under the theory that school-based health centers decrease suspension rates by improving adolescent mental health, we would expect the decrease in suspensions to be the greatest for students whose delinquency is caused by behavioral issues. Psychology research suggests that one common way that behavioral and mental health issues manifest for adolescents, is through "disruptive" or "aggressive behavior" (Garland et al., 2010); therefore, the fraction of suspensions due to disruptive or aggressive behavior may be one possible proxy for the share of suspensions that are caused by mental health issues. The California Department of Education defines six categories of offenses that can lead to a suspension: violent incidents (with injury), violent incidents (no injury), weapons possession, illicit drug possession or sale, defiance-only incidents, and other offenses.<sup>29</sup> To isolate suspensions resulting from mental health issues, I focus on the category of "defiance-only" suspensions, which is defined as "any suspension associated with a student in which the only offense committed by a student is disruption".<sup>30</sup> The outcome I examine for defiance

timate should be interpreted as the average treatment effect across all schools that opened an SBHC in any year between 2012-2019.

 $<sup>^{27}</sup>$ Appendix Table A.11 shows the difference-in-differences estimates by gradespan. The decrease in suspensions is primarily concentrated amongst the middle and high schools in the sample, which is unsurprising given that suspension is fairly uncommon at the elementary school level.

<sup>&</sup>lt;sup>28</sup>Specifically, this specification restricts to the subsample of SBHCs that report having "mental health" services and the matched control schools for those SBHCs. 38 out of 44 SBHCs in this sample report having mental health services.

 $<sup>^{29}</sup>$ The specific offenses included in each of these categories are outlined in Appendix Section B.

<sup>&</sup>lt;sup>30</sup>It is worth noting that violence and drug use are behaviors that have also been linked to mental health issues;

suspensions is the fraction of *total students enrolled* who were suspended for defiance reasons. The corresponding outcome for non-defiance suspensions is the fraction of total students enrolled who were suspended for any reason that does not qualify as "defiance only".

Table 10 shows the estimated event study coefficients for overall suspension rates (Column 1), the defiance-only suspension rate (Column 2), and the non-defiance suspension rate (Column 3). In both Columns (2) and (3) there is no evidence of a significant pre-trend or post-trend; however the coefficients for defiance suspensions in Column (3) are notably larger in magnitude than the coefficients for non-defiance suspensions in Column (2). This is especially noteworthy given that the at baseline, defiance-only suspensions compose a *smaller* fraction of the overall suspension rate compared to non-defiance suspensions (one-third versus two-thirds). The difference in trends for non-defiance versus defiance suspensions is more visible in Figure 4, which plots the treatment coefficients in each time period for non-defiance suspensions (on the left) and defiance suspensions (on the right). While neither of the two offense types shows a statistically significant decrease, there is a visible declining trend for defiance suspensions that suggests that the overall decrease in suspensions is more likely to be driven by defiance than non-defiance suspensions.

Table 11 shows difference-in-differences estimates for the overall, defiance-only, and non-defiance suspension rates. Although there is no significant treatment effect for either non-defiance suspensions (Column 2) or defiance suspensions (Column 3), the relative magnitudes suggest that defiance suspensions compose a larger portion of the decrease in overall suspensions, despite having a lower baseline rate in the control group sample.

These results provide suggestive (but statistically insignificant) evidence that the decrease in suspensions could be driven by the treatment of behavioral and mental health disorders that would lead to disruptive behavior. There are two potential mechanisms for this decrease that we might consider. The first is that the opening of an SBHC improves access to mental health services, thus increasing the fraction of students with behavioral issues who receive treatment and decreasing the frequency of behaviors that would warrant suspension. The second is that the opening of an SBHC leads teachers and principals to refer students with behavioral issues to the mental health professional at the SBHC as an *alternate policy* to suspensions. These mechanisms are not mutually exclusive and under certain conditions may achieve the same impact. Consider, as an illustrative example, a student with untreated ADHD that causes inattention and leads them to act out in class. Assume that prior to the opening of an SBHC, after repeated incidents of acting out in class this student would have been suspended. Under the first mechanism, after the SBHC opens, this student may seek out psychological services from the SBHC and receive the relevant treatments (either in the form of medication or counseling) to improve their focus and prevent disruptive behaviors that would have led to suspension. Under the second mechanism, after the SBHC opens the student's would have led to suspension.

however these are issues that may result from more severe issues that affect a smaller fraction of students. I focus on disruptive behavior as a proxy for mental health issues that may be less severe but still detrimental for students' outcomes if left un-treated. These less severe mental health issues may also be the ones that School-Based Health Centers are best-equipped to treat directly.

	(1)	(0)	(2)
	(1)	(Z)	(3)
	All	Non-Defiance	Defiance
Treated x ( $\tau = -3$ )	-0.0017	0.0010	-0.0038
	(0.0080)	(0.0060)	(0.0061)
Treated x ( $\tau = -2$ )	-0.0063	-0.0057	-0.0014
	(0.0060)	(0.0045)	(0.0049)
Treated x ( $\tau = -1$ )	ref.	ref.	ref.
Treated x ( $\tau = 0$ )	-0.0085	-0.0028	-0.0068
	(0.0053)	(0.0051)	(0.0055)
Treated x ( $\tau = 1$ )	-0.012	-0.004	-0.009
	(0.008)	(0.004)	(0.009)
Treated x ( $\tau = 2$ )	-0.014	0.001	-0.013
	(0.010)	(0.005)	(0.011)
Treated x ( $\tau = 3$ )	-0.015	0.005	-0.013
	(0.010)	(0.008)	(0.013)
F-Stat/Chi-Stat	3.090	2.098	3.004
p-value	0.543	0.718	0.557
Pre-Period Control Mean	0.065	0.044	0.023
Observations	867	812	867

 Table 10:
 Event Study: Suspensions by Offense-Types

Standard errors in parentheses. Observations are at the school level.

F-stat and p-value come from a test that the coefficients on

Treatment X Event-Time for all post-event years are jointly equal to 0. \* p<0.1, \*\* p<0.05, \*\*\* p<0.01

Table 11: Suspensions Rates: Heterogeneity by Offense Type (DiD)

	(1)	(2)	(3)
	All	Non-Defiance	Defiance
Treated X Post	-0.0136**	-0.0051	-0.0086
	(0.0059)	(0.0031)	(0.0062)
Pre-Period Control Mean	0.0649	0.0420	0.0230
Observations	867	867	867

Standard errors in parentheses. Observations are at the school level.

\* p<0.1, \*\* p<0.05, \*\*\* p<0.01



Note: Shaded area represents the 95% confidence intervals

Figure 4: This figure plots the *Event Time* coefficients from separate event studies where the outcomes are: the suspension rate for defiance suspensions (left) and the suspension rate for non-defiance suspensions (right). Both sub-sample event studies control for school fixed effects and a vector of school characteristics that includes fraction of Free and Reduced Price Meal (FRPM) students, fraction of minority students, and total school enrollment. All lags prior to event time -3 and all leads after event time 3 are dropped from the estimation sample. Standard errors are clustered at the school level.

teacher chooses to refer them to see the SBHC's psychologist after multiple disruptive incidents, rather than sending them to the principal to receive a disciplinary repercussion. If the referral to receive mental health services ultimately leads the student to receive treatment for the issues that lead to their disruptive behavior, then the second mechanism has an equivalent outcome to the first mechanism. If, on the other hand, the student is referred to mental health services as an alternative to suspension but never actually receives treatment or changes their behavior, then a decrease in their likelihood of suspension would not be indicative of an improvement in their mental health.

One approach to disentangling these two mechanisms is to examine the effect of SBHCs on the rate of repeat suspensions. Specifically, I consider the number of suspensions per suspended student as a measure of the rate of repeat suspensions. If SBHCs do not actually decrease behavioral issues and only replace suspensions for behavioral issues with referrals to a mental health professional, we might expect that this "policy" will affect the extensive margin (i.e. whether or not any given student is suspended) rather than the intensive margin (i.e. how many times a student is suspended for disruptive behavior. If, on the other hand, SBHCs are reducing disruptive behavior by improving mental health for certain students, this should both reduce the extensive marginal probability of suspension for some subset of students, and decrease the likelihood of repeated suspensions for those students who have already been suspended at least once. Figure 5 shows the results from an event study regression where the outcome is the school-level average number of suspensions per suspended student. As with the regressions for the overall suspension rate, I find no effects of the treatment in the pre-SBHC period. There is, however, a downward trend beginning in the year of the SBHC opening. The effect of the treatment is statistically significant one year following treatment, and insignificant in the following years. Appendix tables A.6 and A.7 show the corresponding event study and difference-in-differences estimates. While the difference-in-difference estimates are not statistically significant, in combination with the visible declining trend in the event study figure they provide weak evidence that the number of suspensions per suspended student may decrease by around 0.09 suspensions in the years following an SBHC opening.

Figure 6 decomposes the change in suspensions per suspended student by defiance-only and non-defiance suspensions. Although both figures are noisy and difficult to interpret, the righthand figure does suggest a clearer "declining" trend in the number of defiance suspensions per suspended student in the years following an SBHC. The trend for non-defiance suspensions is notably less clear. It is worth interpreting these trends cautiously, since the corresponding difference-in-differences estimates in Appendix table A.8 suggest that the average effect in the post-period is of similar magnitude for defiance and non-defiance suspensions. However the trends do suggest that the patterns identified in the overall suspension rate may also be present in the repeat suspension rate.<sup>31</sup> The decreases on both the extensive and intensive margin captured in these results is consistent

 $<sup>^{31}</sup>$ Appendix Table A.5 shows the difference-in-differences results from a similar analysis for the total number of suspensions per student across the entire student population of a school. This alternate metric captures the changes on both the extensive and intensive margins. Unsurprisingly given the previous results, I find that the number of suspensions per student decreases by around 55%, from the control baseline of 0.07 suspensions per student.



Figure 5: This figure plots the *Treatment*  $\times$  *Event Time* coefficients from an augmented event study where the outcome is the number of suspensions per suspended student. The regression depicted controls for school fixed effects and a vector of school characteristics that includes fraction of Free and Reduced Price Meal (FRPM) students, fraction of underrepresented minority students, and total school enrollment. All lags prior to event time -3 and all leads after event time 3 are dropped from the estimation sample. Standard errors are clustered at the school level.

with the theory that part of the decrease in suspensions due to SBHC-access can be explained by a decrease in repeated behavioral issues.



Suspensions per Suspended Student: Defiance vs Non-Defiance Event Studies

Note: Shaded area represents the 95% confidence intervals

Figure 6: This figure plots the *Treatment*  $\times$  *Event Time* coefficients from three event studies where the outcome is the number of suspensions per suspended student. From left to right, the relevant samples are: all suspensions, defiance-only suspensions, and non-defiance suspension. All three specifications control for school fixed effects and a vector of school characteristics that includes fraction of Free and Reduced Price Meal (FRPM) students, fraction of underrepresented minority students, and total school enrollment. All lags prior to event time -3 and all leads after event time 3 are dropped from the estimation sample. Standard errors are clustered at the school level.

#### 5.3 Dropout Rates

Following a similar structure to the previous section, Table 12 shows the results from event-study specifications where the dependent variable is now the school-level dropout rate.<sup>32</sup> It is worth noting that since dropouts most commonly occur at the high school level, the sample of schools included in this analysis includes predominantly high schools, and a few middle schools if they report dropout rates.<sup>33</sup> As a result, there is insufficient sample to estimate a robustness check specification that includes gradespan time trends. Appendix Figure A.8 shows that my main results do not change when restricting only to high schools in the sample. As before, Column (2) shows the results from my preferred specification. Notably, since the available data on dropout rates is more limited than the available data on suspension rates these analyses are restricted to smaller sample sizes, which poses additional concerns for power.

Table 12 shows no treatment effects in the pre-opening years, which once again suggests that the parallel trends assumption is satisfied. While there is no significant impact of opening an SBHC on dropout rates for up to two years following the opening, there is a decrease of around 1.2 percentage points in year 3 that is statistically significant at the 5% level. Compared to the control school baseline dropout rate of 0.8% this is a nearly 150% decrease. It is worth noting that the Year 3 coefficient is estimated on a relatively small sample of schools and therefore should be interpreted with caution. Table 13, which contains the corresponding difference-in-differences estimate, suggests that the average effect in the post-period is close to zero, suggesting that the drop in year 3 is more likely to be spurious than meaningful. Finally, figure A.7 shows the corresponding event study plot, once again confirming that there is a close to zero effect both in the pre-period and post-period. Finally, Table 13 shows the corresponding difference-in-differences estimates, which show an average treatment effect of around 0.1 percentage points that is insignificant with the addition of controls in Columns (2) and (3).

A zero-effect is in line with previous papers in this literature. Lovenheim et al. (2016) similarly finds no identifiable effect on dropout rates, indicating that a null effect should not necessarily be shocking. There is still, however, value to considering what effect sizes could be ruled out by these results. In particular, if dropout rates increase after an SBHC opening, we may be worried that the estimated decrease in suspension rates is due to a "crowd out" effect, where students that may otherwise have been suspended are now instead dropping out. The 95% confidence intervals on the difference-in-differences estimates rule out increases or decreases in suspension rates is around 1.1 percentage points, this suggests that an *increase in dropout rates* is unlikely to fully explain the decrease in suspension rates.

 $<sup>^{32}\</sup>mathrm{The}$  corresponding figure is shown in Appendix Figure A.7

 $<sup>^{33}</sup>$ Appendix Table A.12 shows the primary difference-in-differences estimates separately by gradespan. There are no differential effects when looking only at high schools or middle schools.

	(1)	(2)
	Baseline	Demographics
Treated x ( $\tau = -3$ )	0.0006	0.0007
	(0.0025)	(0.0025)
Treated x ( $\tau = -2$ )	0.0018	0.0021
	(0.0028)	(0.0028)
Treated x ( $\tau = -1$ )	ref.	ref.
Treated x ( $\tau = 0$ )	-0.0011	0.0009
	(0.0016)	(0.0023)
Treated x ( $\tau = 1$ )	-0.001	0.000
	(0.001)	(0.002)
Treated x ( $\tau = 2$ )	-0.004	-0.003
	(0.003)	(0.002)
Treated x ( $\tau = 3$ )	-0.016***	-0.012***
	(0.003)	(0.003)
F-Stat/Chi-Stat	37.173	24.087
p-value	0.000	0.000
Pre-Period Control Mean	0.008	0.008
Observations	372	372

Table 12: Dropout Rates: Event Study Specifications

Standard errors in parentheses. Observations are at the school level. F-stat and p-value come from a test that the coefficients on Treatment X Event-Time for all post-event years are jointly equal to 0. \* p<0.1, \*\* p<0.05, \*\*\* p<0.01

Table 13: Dropout Rates: Difference-in-Differences Specifications

	(1)	(2)
	Baseline	Demographics
Treated X Post	-0.0020**	-0.0012
	(0.0010)	(0.0010)
Pre-Period Control Mean	0.0084	0.0084
$R^2$	372	372

Standard errors in parentheses. Observations are at the school level. \* p<0.1, \*\* p<0.05, \*\*\* p<0.01
### 6 Robustness

#### 6.1 Staggered Implementation Adjustments

A concern with difference-in-difference models in staggered adoption settings is the potential for biased estimation of average treatment effects in the presence of treatment effect heterogeneity across treated units or time periods. A recent and growing literature shows that use of a simple two-way fixed effects model in the context of staggered treatment adoption can lead to incorrect average treatment effects (ATEs), that in an extreme case, may of an opposite sign to the true treatment effects for each treated group (Goodman-Bacon (2021) Callaway and Sant'Anna (2021), Sun and Abraham (2021), and de Chaisemartin and D'Haultfœuille (2020)). Goodman-Bacon (2021) shows that a difference-in-differences estimate with multiple treatment years can be decomposed into a weighted average of every possible sub-sample difference-in-differences, that compares a given treated group and control group before and after the treatment time. The potential for bias in this weighted average comes from the possibility of negative weights assigned to certain comparisons; in particular, the use of *already treated* units as controls for future treated units may lead to negative weighting for the early-treated units. Papers such as Sun and Abraham (2021) have shown that parallel concerns arise with event study models.

The propensity score matching approach employed in this paper takes significant steps to address this concern. For one, the matched control group consists of only *never treated schools*, preempting the issues with accidentally using already treated units as controls for later-treated units. This does not, however, fully obviate the possibility of negative weighting. Since a difference-in-differences estimate compares the average outcome for all treated schools after their respective treatment years to all observations that are untreated in prior years, the unit and time fixed effects are still likely to be estimated, in part, by already treated units. Several papers have proposed methods to test for bias from treatment effect heterogeneity and alternate estimators that compensate for the possibility of negative weighting. The primary specifications in Section 5come from the alternate estimator proposed by Callaway and Sant'Anna (2021). The Callaway and Sant'Anna estimator addresses bias by first calculating a two-period difference-in-differences estimate for each treatment cohort (i.e. group of schools treated in the same year) compared to all never-treated observations, between any two years y and z. These two-period estimated ATTs can then be aggregated into weighted averages at the event-time (i.e. time relative to the treatment event), group (i.e. treated cohort), time period (i.e. year), or overall sample level using a weighting method selected by the researchers.<sup>34</sup>

In this section I attempt to further bolster the claim that bias from heterogeneous treatment

<sup>&</sup>lt;sup>34</sup>Following recommendations from Callaway and Sant'Anna (2021), I use doubly-robust inverse propensity weighting, which weights each estimate proportionally to the size of the treated cohort. Roth et al. (2023) notes that this is an improvement on the standard two-way fixed effects estimator, which allows the OLS process to determine the weights, thereby leading to each group being weighted proportionally to the variance of the treatment dummy.

effects is not a major concern in my sample. Following the procedure proposed in de Chaisemartin and D'Haultfœuille (2020), I calculate the weights on each decomposed difference-in-difference estimate and use the standard deviation of the weights to estimate a bound on the standard deviation of treatment effects (which de Chaisemartin and D'Haultfœuille define as the level of "treatment effect heterogeneity") that would be necessary: (a) for the true treatment effect to be 0; and (b) for all sub-group average treatment effects to have a *different sign* from the estimated treatment effect for the full sample.<sup>35</sup> For my preferred specification, 95% of weights are positive. The necessary level of heterogeneity is bounded at 0.02 for the true treatment effect to be 0, and 0.017 for the true treatment effect to be positive. These bounds suggest that there would need to be a fairly high amount of treatment effect heterogeneity (equivalent to nearly 46% of the baseline suspension rate) for the true ATE to be of the opposite sign as the estimated ATE. Based on these bounds and the high fraction of positive weights, there seems to be minimal concern that the estimated treatment effects from the primary models are severely biased by treatment-effect heterogeneity. The high fraction of positive weights and the bounds derived from the standard deviation of those weights suggest that at minimum we can be confident that if there is a true effect of SBHC openings on suspensions, it is most likely a decrease. The Callaway and Sant'Anna estimates shown in Section 5 confirm that correcting for this bias does not meaningfully alter the magnitude and significance of the treatment effects.

## 7 Heterogeneity Analyses

Given the large magnitude decreases in suspension rates identified in the previous section, it is of interest to consider which groups and mechanisms may be driving these effects. I assess this by looking at heterogeneity of treatment effects by gender. The CDE data on suspension rates provides a novel opportunity for this heterogeneity analysis as it provides suspension rates and counts for each school disaggregated by gender, race, and category of offense. All event study estimates follow the model from Equation 1 and all difference-in-differences estimates follow Equation 2. Once again, the preferred specification includes school fixed effects and controls for a vector of school characteristics most likely to be correlated with opening an SBHC.

Examining heterogeneity by gender, Figure 7 shows separate event study regressions for the sub-samples of male and female students. The outcome for males should be interpreted as the fraction of male students in a school that were suspended in a year; similarly, the outcome for females is the fraction of female students in a school suspended in a year. The primary takeaway from this figure is that the decrease in suspension rates identified in the previous section seems to be primarily driven by a decline in suspension rates for male students. This is in line with a much of the literature on delinquency; Komisarow and Hemelt (2022), for example, finds that the effects of their telemedicine program on chronic absenteeism and delinquency are greatest for male students.

 $<sup>^{35}</sup>$ In practice, this is implemented using the *twowayfeweights* package in Stata.

Appendix Table A.13 show the corresponding event study estimates that support the presence of this heterogeneity.



Suspension Rates: Event Studies by Gender

Note: Shaded area represents the 95% confidence intervals

Figure 7: This figure plots the *Event Time* coefficients from separate event studies for the outcomes of male suspension rate (left) and female suspension rates (right). Both sub-sample event studies control for school fixed effects and a vector of school characteristics that includes fraction of Free and Reduced Price Meal (FRPM) students, fraction of underrepresented minority students, and total school enrollment. All lags prior to event time -3 and all leads after event time 3 are dropped from the estimation sample. Standard errors are clustered at the school level.

Finally, table 14 shows difference-in-differences estimates for the same three samples. Now considering the average change in suspension rates after the opening, the decrease in suspension rates for male students is around 2 percentage points, while the corresponding decrease for female students in only 0.6 percentage points. This suggests that while both male and female students are positively affected by the opening of an SBHC, the impact may be stronger for males.

	М	ale Suspension F	Rate	Female Suspension Rate			
	(1)	(1) $(2)$ $(3)$		(4)	(5)	(6)	
	Baseline	Demographics	$\operatorname{Gradespan}$	Baseline	Demographics	$\operatorname{Gradespan}$	
Treated X Post	-0.0190**	-0.0206**	-0.0249**	-0.0067	-0.0064	-0.0090	
	(0.0085)	(0.0089)	(0.0110)	(0.0041)	(0.0041)	(0.0056)	
Pre-Period Control Mean	0.0872	0.0872	0.0897	0.0410	0.0410	0.0421	
Observations	867	867	812	867	867	812	

Table 14: Suspensions Rates: Gender Heterogeneity (DiD)

Standard errors in parentheses. Observations are at the school level. \* p<0.1, \*\* p<0.05, \*\*\* p<0.01

## 8 Discussion

The results of this analysis show that access to a school-based health center leads to a significant drop in suspension rates. I also present suggestive (but weaker) evidence that this drop may be primarily driven by "defiance" suspensions. The decomposition of the effect on suspension rates, in combination with descriptive evidence that higher suspensions are positively correlated with worse mental health, suggest that the drop in suspension rates could be driven by an *improvement* in students' mental health which in turn reduces disruptive behavior. In order to assess the policy implications of these results, it is important to consider the alternative channels through which the opening of a school-based health center might lead to a drop in suspension rates. One alternative mechanism that this paper is unable to rule out is that SBHCs may be *displacing* disciplinary approaches to addressing disruptive behavior. Anecdotal evidence from administrators at the California School-Based Health Alliance suggests that school-based health centers strive to work closely with classroom teachers and school administrators to ensure that their services are integrated into the broader system of support the school offers its students. If this integration leads school administrators to send a disruptive student to the SBHC as an *alternative* to suspending them, then the opening of an SBHC may lead to a decline in suspensions independently of whether it causes improvement in students' mental health.

There are contextual reasons that this "displacement" channel is unlikely to fully explain the decline in suspension rates. For example, discussions with SBHC administrators have indicated that that more often than not, SBHCs are insufficiently staffed relative to the demand for mental health services. The strain on these services should disincentivize school principals from sending students to SBHCs unless they deem that the student would actually receive and benefit from treatment. However, even if the decrease in suspensions is primarily driven by displacement, this does not necessarily suggest that SBHCs have no impact on adolescent mental health. For many students, that initial visit with a mental health professional may have a positive effect on helping them identify their own mental health treatment needs or overcome the baseline stigma they may hold regarding therapeutic services. With these considerations in mind, future work should focus on disentangling these mechanisms through conversations with SBHC leaders and school administrators, and student-

level data on SBHC utilization. Understanding the exact mechanisms driving this decrease is critical for assessing the direct mental health impacts of SBHCs. However, regardless of which mechanisms are driving these effects, one potential takeaway for policymakers is that the provision of in-school mental health services may be a valuable approach to decreasing students' exposure to discipline.

A second important takeaway for policymakers is that the school-based health center model has high potential impacts for low-income communities, but the impact of expanding these services to higher-income and less racially diverse communities is inconclusive. The process of identifying a theoretically and empirically appropriate control group leads to results that are local to schools that serve a larger number of students and have a higher fraction of free-and-reduced price lunch students. This paper is unable to conclude that school-based health centers would have the same magnitude impact on schools that look meaningfully different from the treated and control samples. Exploring the heterogeneous impacts of SBHCs on schools with different demographic profiles requires a larger sample of SBHCs than are available in this study, and is a valuable goal for future research.

Finally, while this paper provides important evidence on an outcome that is strongly correlated with worse mental health, data limitations prevent me from identifying the direct effect of these centers on students' mental health. If access to an SBHC decreases suspensions through improved mental health, the total benefits accrued to students would include the direct benefits from improved mental health, the direct benefit from decreased discipline, the indirect benefits that those intermediate outcomes have on long-run academic achievement and labor-market outcomes, and any spillovers to students who are not directly utilizing SBHC services. Since my analysis is at the school-level, assessing students' long-run labor market outcomes is infeasible; therefore, translating short-run effects into long-run impacts would require a causal pathway from decreased suspension rates to long-run earnings and labor market participation.

The evidence on these causal pathways is limited and difficult to translate to my setting due to contextual differences. However, recent papers have suggested that higher suspension rates may have long-run impacts on student performance, dropout rates, and future incarceration. Descriptive work from Fabelo et al. (2011) shows that being suspended is positively correlated with repeating a grade and dropping out of school. Recent quasi-experimental work from Bacher-Hicks et al. (2019) finds that students who are randomly assigned to attend a middle school with a one standard deviation higher suspension rate than their previous school are 1.7 percentage points more likely to ever drop out of school and 2.5 percentage points more likely to ever be incarcerated. These results suggest that by decreasing suspension rates, school-based health centers may also have longer-run impacts on students' high school completion and employment. Finally, work from Carrell et al. (2018) shows that exposure to a disruptive peer in elementary school has long-run effects on reducing earnings by around 3%. If one of the mechanisms for the drop in suspension rates is a true decrease in disruptive behavior, there may be positive spillovers from decreases in classroom disruptions for students who are not directly using SBHC services that are difficult to quantify in my analysis. Developing policy recommendations regarding the cost-effectiveness of school-based health centers

requires further research on the direct effects of these centers on student mental health and long-run outcomes, to supplement the effects identified in this paper.

# 9 Conclusion

Worsening trends in adolescent mental health have been a focal point of recent policy discussions and funding investments. This paper aims to contribute to that policy discussion by evaluating the impact of access to a school-based health center on suspension and dropout rates, two behavioral outcomes that are likely to be directly impacted by untreated mental health issues. In addition to being the first paper to examine the effect of school-based health center access on suspensions and the first to examine the impact on dropout rates using school-level data, this paper also provides novel evidence regarding the correlation between these outcomes and student-reported mental health and school climate. To address non-randomness in the decision to open an SBHC, I leverage a propensity-score matching approach to identify a theoretically reasonable control group and show that the trends in outcomes look similar between treated and control schools in the years leading up to an opening.

I find evidence that the opening of an on-site SBHC decreases suspension rates by 1.3 percentage points when compared the propensity-matched schools. To put this drop in suspension rates into perspective, the baseline average suspension rate for control schools is only 6.5%, so this predicted range of effects represents a 20% decrease in suspension rates. Examining the mechanisms for the change in suspension rates, I provide suggestive evidence that the decrease may be driven by an improvement in mental health. First, correlations from the California Healthy Kids Survey suggest that high suspension rates are negatively correlated with feelings of "belonging" and "staff support" and positively correlated with feelings of depression. Decomposing suspension rates by the category of offense, I find a decreasing trend in suspensions for disruptive behavior, which is often the result of untreated mental health issues. Finally, I show suggestive evidence that the decrease in overall suspension rates may be complemented by a decrease in the rate of repeat suspensions. indicating that the opening of SBHCs might actually change behaviors that lead to suspension rather than simply changing a school's policies around dealing with disruptive students. While the latter two results are not statistically significant (likely due to a small sample of treated schools), the combination of results points to potential positive impacts of school-based health centers on mental health issues that would manifest in behavioral issues if left untreated. Future work should aim to explore this channel further, ideally with a larger sample of schools to increase statistical power.

While I find no overall effect of SBHC-access on dropout rates, tight confidence intervals on these estimates allow me to rule out increases and decreases in dropout rates of more than 0.5 of a percentage points. This indicates that the decrease in suspension rates is unlikely to be the result of *crowd-out* by an increase in dropout rates of similar magnitude. Specifically, this helps rule out the possibility that SBHCs lead students who would otherwise have been suspended, to drop out of school instead. It is also helpful to note that a zero-effect for dropout rates is in line with the results of previous research on school-based health centers by Lovenheim et al. (2016).

There are a number of reasons why school-based health centers may not directly affect dropout rates. For one, the correlation between mental health and dropout rates is more ambiguous than the correlation between mental health and suspension rates, suggesting that if school-based health centers have an impact on mental health, this would be more likely to be captured by changes in suspensions than changes in dropout rates. This is not unreasonable given that the long-run repercussions to dropping out are larger than the repercussions of a single suspension. Moreover, while the decision to drop out could be linked to poor mental health, alternative factors such as family issues and academic performance may be stronger drivers. Finally, it is worth noting that this study is unable to rule out longer-run effects on dropout rates that show up more than three years after the SBHC opening.

The results I present must be cautiously interpreted relative to the assumptions made in selecting the control group. Specifically, we can conclude that suspension rates decrease in the years following the opening of an SBHC for treated schools relative to untreated schools in districts that have opened an SBHC prior to the current opening. The effect that is isolated can be interpreted as the effect of having an SBHC in a specific school on the students of that school compared to students from a demographically similar school in a district with a similar underlying propensity to open an SBHC. I run several alternate specifications to test the robustness of this model and find that expanding from one-to-one to two-to-one matching strengthens the significance of the treatment effects. However, these tests also reveal that the primary results may not be robust to alternate choices of control group. While there are contextual reasons that each alternate control group may be a theoretically "bad" match, the limitations of my data and identification strategy make it difficult to rule out these control groups entirely. Therefore, while the results from my primary specifications are highly suggestive of positive impacts of school-based health centers on outcomes linked to mental health improvement, these results may not replicate in districts that are significantly different from the treated districts in my sample. In part, this study highlights the challenges to causal identification using existing data on SBHCs. Given these challenges, rigorous evaluation of these centers would benefit greatly from a more exogenous source of variation such as a policy that randomizes the timing at which new schools receive grants to fund new SBHC construction.

This paper contributes to a small but growing literature on the impacts of school-based health centers. My results provide novel evidence that these centers may have large impacts on student behavior, and suggest a few natural paths for future research. First, developing a clear recommendation on the use of SBHCs to address adolescent mental health requires evidence on the direct mental-health impacts of these centers. Moreover, since there are large overhead costs to opening and operating SBHCs, it may be valuable to consider whether less comprehensive or intensive alternatives to on-site SBHCs, such as in-school mental health professionals or mobile SBHC clinics, could have the same behavioral impacts. Finally, in order to better assess how much variation there is the impact of SBHCs, one research priority should be to collect and standardize data on individual SBHC operations, services, and utilization. This data would improve our understanding of the specific features of SBHCs that benefit students, and provide valuable insight into how the effects of these centers can be replicated across different settings.

## References

- Bacher-Hicks, A., Billings, S. B., and Deming, D. J. (2019). The school to prison pipeline: Long-run impacts of school suspensions on adult crime. Technical report, National Bureau of Economic Research.
- BHWF (2023). Primary care health professional shortage areas (hpsas).
- Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. Journal of econometrics, 225(2):200–230.
- Carrell, S. E. and Carrell, S. A. (2006). Do lower student to counselor ratios reduce school disciplinary problems? The BE Journal of Economic Analysis & Policy, 5(1):0000101515153806451463.
- Carrell, S. E. and Hoekstra, M. (2014). Are school counselors an effective education input? *Economics Letters*, 125(1):66–69.
- Carrell, S. E., Hoekstra, M., and Kuka, E. (2018). The long-run effects of disruptive peers. American Economic Review, 108(11):3377–3415.
- CDC (2022). New cdc data illuminate youth mental health threats during the covid-19 pandemic.
- CDE (2023). Income eligibility scales for school year 2023–24.
- CSBHA (2022a). From vision to reality. Technical report, California School-Based Health Alliance, Oakland, CA.
- CSBHA (2022b). Navigating the promise of sbhcs: A guide for health care leaders.
- CSBHA (2023). Funding school-based health & wellness centers: California school-based health alliance.
- CSHCA (2010). California school nursing roles and responsibilities.
- Currie, J. and Stabile, M. (2006). Child mental health and human capital accumulation: the case of adhd. *Journal of health economics*, 25(6):1094–1118.
- de Chaisemartin, C. and D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96.
- Deshpande, M. and Li, Y. (2019). Who is screened out? application costs and the targeting of disability programs. *American Economic Journal: Economic Policy*, 11(4):213–48.
- Fabelo, T., Thompson, M. D., Plotkin, M., Carmichael, D., Marchbanks, M. P., and Booth, E. A. (2011). Breaking schools' rules: A statewide study of how school discipline relates to students' success and juvenile justice involvement. New York: Council of State Governments Justice Center.

- Fadlon, I. and Nielsen, T. H. (2021). Family labor supply responses to severe health shocks: Evidence from danish administrative records. American Economic Journal: Applied Economics, 13(3):1– 30.
- Flaherty, L. T. and Osher, D. (2003). History of school-based mental health services in the united states. In *Handbook of school mental health advancing practice and research*, pages 11–22. Springer.
- Gall, G., Pagano, M. E., Desmond, M. S., Perrin, J. M., and Murphy, J. M. (2000). Utility of psychosocial screening at a school-based health center. *Journal of School Health*, 70(7):292–298.
- Garland, A. F., Brookman-Frazee, L., Hurlburt, M. S., Accurso, E. C., Zoffness, R. J., Haine-Schlagel, R., and Ganger, W. (2010). Mental health care for children with disruptive behavior problems: A view inside therapists' offices. *Psychiatric Services*, 61(8):788–795.
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal* of *Econometrics*, 225(2):254–277.
- Hanson, T. L. and Kim, J.-O. (2007). Measuring resilience and youth development: the psychometric properties of the healthy kids survey.
- Heckman, J. J., Ichimura, H., and Todd, P. (1998). Matching as an econometric evaluation estimator. The review of economic studies, 65(2):261–294.
- Heckman, J. J., Ichimura, H., and Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The review of economic studies*, 64(4):605–654.
- Juszczak, L., Melinkovich, P., and Kaplan, D. (2003). Use of health and mental health services by adolescents across multiple delivery sites. *Journal of Adolescent Health*, 32(6):108–118.
- Kerns, S. E., Pullmann, M. D., Walker, S. C., Lyon, A. R., Cosgrove, T., and Bruns, E. J. (2011). Adolescent use of school-based health centers and high school dropout. Archives of Pediatrics & Adolescent Medicine, 165(7):617–623.
- Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., and Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of dsm-iv disorders in the national comorbidity survey replication. Archives of general psychiatry, 62(6):593–602.
- Kisker, E. E. and Brown, R. S. (1996). Do school-based health centers improve adolescents' access to health care, health status, and risk-taking behavior? *Journal of Adolescent Health*, 18(5):335–343.
- Komisarow, S. and Hemelt, S. W. (2022). School-based healthcare and absenteeism: Evidence from telemedicine. *Education Finance and Policy*, pages 1–50.

- Love, H. E., Schlitt, J., Soleimanpour, S., Panchal, N., and Behr, C. (2019). Twenty years of school-based health care growth and expansion. *Health Affairs*, 38(5):755–764.
- Lovenheim, M. F., Reback, R., and Wedenoja, L. (2016). How does access to health care affect teen fertility and high school dropout rates? evidence from school-based health centers. Technical report, National Bureau of Economic Research.
- Mahecha, J. and Hanson, T. (2020). Measurement structure of the california school climate, health, and learning surveys.
- McCord, M. T., Klein, J. D., Foy, J. M., and Fothergill, K. (1993). School-based clinic use and school performance. *Journal of Adolescent Health*, 14(2):91–98.
- McLeod, J. D., Uemura, R., and Rohrman, S. (2012). Adolescent mental health, behavior problems, and academic achievement. *Journal of health and social behavior*, 53(4):482–497.
- NCBH (2019). America's mental health.(2018).
- Paschall, M. J. and Bersamin, M. (2018). School-based health centers, depression, and suicide risk among adolescents. American journal of preventive medicine, 54(1):44–50.
- Reback, R. (2010a). Noninstructional spending improves noncognitive outcomes: Discontinuity evidence from a unique elementary school counselor financing system. *Education Finance and Policy*, 5(2):105–137.
- Reback, R. (2010b). Schools' mental health services and young children's emotions, behavior, and learning. *Journal of Policy Analysis and Management*, 29(4):698–725.
- Roth, J., Sant'Anna, P. H., Bilinski, A., and Poe, J. (2023). What's trending in difference-indifferences? a synthesis of the recent econometrics literature. *Journal of Econometrics*.
- Rott, N. (2013). La schools throw out suspensions for 'willful defiance'.
- SAMHSA (2023). Highlights for the 2021 national survey on drug use and health.
- Santelli, J., Kouzis, A., and Newcomer, S. (1996). Student attitudes toward school-based health centers. Journal of Adolescent Health, 18(5):349–356.
- Smith, J. A. and Todd, P. E. (2005). Does matching overcome lalonde's critique of nonexperimental estimators? *Journal of econometrics*, 125(1-2):305–353.
- Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.
- Walker, S. C., Kerns, S. E., Lyon, A. R., Bruns, E. J., and Cosgrove, T. (2010). Impact of schoolbased health center use on academic outcomes. *Journal of Adolescent Health*, 46(3):251–257.

- Warren, C. and Fancsali, C. (2000). Lessons from the evaluation of new jersey's school-based youth services program. Improving Results for Children and Families: Linking Collaborative Services with School Reform Efforts, pages 91–120.
- Wright, B., Padilla-Frausto, D. I., Tse, H. W., Crawford-Roberts, A., Kabir, F., and Salem, S. (2021). Nearly 1 in 3 adolescents in california reports serious psychological distress. Technical report, Los Angeles, Calif.: UCLA Center for Health Policy Research.

# A Additional Tables & Figures

\_

	Untreated	Treated	p-value <sup>†</sup>
Suspension Rate	0.03	0.08	0.000
	[20493]	[198]	
Female Suspension Rate	0.02	0.06	0.000
	[20311]	[198]	
Male Suspension Rate	0.04	0.11	0.000
	[20418]	[198]	
Defiance-Only Suspension Rate	0.01	0.04	0.000
	[20493]	[198]	
Non-defiance Suspension Rate	0.02	0.05	0.000
	[20493]	[198]	
Violence Suspension Rate	0.65	0.51	0.000
	[15906]	[187]	
Weapon Possession Suspension Rate	0.00	0.00	0.000
	[20493]	[198]	
Illicit Drug Suspension Rate	0.01	0.02	0.000
	[20493]	[198]	
Dropout Rate	0.02	0.01	0.261
	[5049]	[142]	
Female Dropout Rate	0.01	0.01	0.273
	[5018]	[142]	
Male Dropout Rate	0.02	0.01	0.324
	[5030]	[142]	

Table A.1: Summary Statistics: Suspension Rates and Dropout Rates (Pre-treatment Only)

 $^{\dagger}$  *p*-values are from a t-test that the treated and un-treated school sample means are equal. Number of observations is listed in brackets under each sample mean.

Table A.2:	Matching	criteria for	nine ite	erations	of fuzz	y string	matching
------------	----------	--------------	----------	----------	---------	----------	----------

Iteration	Matching Criteria
1	Address $SS > 0.917$
2	Matching city, matching school
3	Matching city + matching zip code + highest gradespan SS if school name $SS > 0.45$
4	Matching city + partially matching zip code if school name $SS > 0.45$
5	Matching city + highest school name SS if school name $SS > 0.45$
6	Matching $city + highest zip code SS$
7	Matching city + partial zip match
8	Matching zip
9	Partially matching zip code

SS = Similarity Score. In all rows, "Matching" indicates a similarity score of 1.

Pairs matched in iteration 8 match on the exact zip code.

Pairs matched at iteration 9 have a "partially matching" zipcode, where the zipcode of the school or SBHC

of is nested in the zipcode of the other (eg. SBHC has a zipcode of "95121" and the school has a zipcode of "95121-1845")

#### Table A.3: Correlations Between School Climate and Suspension Rates

	Delinquency		Substance		Caring Staff		School Connectedness	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Year FE	Year/School FE	Year FE	Year/School FE	Year FE	Year/School FE	Year FE	Year/School FE
More Delinquency (1-5)	$0.094^{***}$	0.021***						
	(0.007)	(0.004)						
More Substance Use (1-5)			$0.037^{***}$	$0.007^{***}$				
			(0.005)	(0.003)				
Worse Caring Staff-Student Relationships (1-5)					$0.062^{***}$	$0.009^{***}$		
					(0.003)	(0.003)		
Worse School Connectedness (1-5)							$0.062^{***}$	$0.015^{***}$
							(0.002)	(0.002)
Constant	$0.017^{***}$	$0.070^{***}$	$0.042^{***}$	$0.078^{***}$	$-0.107^{***}$	$0.058^{***}$	-0.065***	0.050***
	(0.006)	(0.004)	(0.006)	(0.004)	(0.008)	(0.009)	(0.005)	(0.006)
Observations	10878.000	10436.000	10881.000	10438.000	10882.000	10441.000	10882.000	10441.000
Sample Mean	0.062	0.063	0.062	0.063	0.062	0.063	0.062	

Standard errors in parentheses. Observations are at the school level. \* p<0.1, \*\* p<0.05, \*\*\* p<0.01

#### Table A.4: Correlations Between School Climate and Dropout Rates

	Delinquency		Substance		Caring Staff		School Connectedness	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Year FE	Year/School FE	Year FE	Year/School FE	Year FE	Year/School FE	Year FE	Year/School FE
More Delinquency (1-5)	0.020***	0.004**						
	(0.002)	(0.002)						
More Substance Use (1-5)			$0.008^{***}$	0.000				
			(0.001)	(0.001)				
Worse Caring Staff-Student Relationships (1-5)					$0.008^{***}$	-0.004***		
					(0.002)	(0.001)		
Worse School Connectedness (1-5)							$0.011^{***}$	-0.003**
							(0.001)	(0.001)
Constant	-0.003	$0.010^{***}$	0.003	$0.013^{***}$	-0.012**	0.025***	-0.014***	0.021***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.005)	(0.004)	(0.003)	(0.004)
Observations	8169.000	7761.000	8172.000	7765.000	8173.000	7767.000	8175.000	7769.000
Sample Mean	0.008	0.008	0.008	0.008	0.008	0.008	0.008	

Standard errors in parentheses. Observations are at the school level. \* p<0.1, \*\* p<0.05, \*\*\* p<0.01



**Figure A.1:** This figure comes from a presentation delivered by the California School-Based Health Alliance (CBSHA) at their annual conference in 2023. It outlines the tiers of mental-health service provision recommended for new SBHCs by the CBSHA. Tier 1 services are services that all SBHCs in California that report offering "mental health" services will provide. The ability to offer Tier 2 and Tier 3 services will vary from center to center and may depend on funding, staffing, and student demand amongst other factors.



**Figure A.2:** This figure shows the event study plots from standard two-way fixed-effects regressions where the outcome is suspension rates. From **left to right** the specifications pictured contain: no demographic controls; a vector of school characteristic controls; and school characteristic controls plus gradespan time trends.

	(1)	(2)	(3)	(4)	(5)	(6)
	Any Offense	Defiance	Non-Defiance (All)	Violence	Weapon Poss.	Illicit Drug
Treated X Post	-0.0099	-0.0075	-0.0024	-0.0105*	0.0002	-0.0018
	(0.0065)	(0.0073)	(0.0039)	(0.0063)	(0.0005)	(0.0019)
Fraction FRPM	0.0128	0.0151	-0.0022	-0.0140	0.0021	-0.0120
	(0.0342)	(0.0319)	(0.0156)	(0.0252)	(0.0017)	(0.0124)
Fraction Minority	0.0873	-0.0459	$0.1332^{***}$	0.1295	0.0046	$0.0408^{**}$
	(0.0625)	(0.0810)	(0.0500)	(0.0879)	(0.0029)	(0.0176)
School Size	-0.000	-0.000	0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.016	0.083	-0.067**	-0.033	-0.004*	-0.019*
	(0.058)	(0.066)	(0.030)	(0.050)	(0.002)	(0.011)
School Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Pre-Period Control Mean	0.0649	0.0230	0.0420	0.0533	0.0028	0.0128
$R^2$	0.807	0.574	0.782	0.763	0.469	0.707
Observations	867	867	867	867	867	867

Table A.5: Suspensions Rates: Heterogeneity by Offense Type (DiD)

Observations are at the school level.

\* p<0.1, \*\* p<0.05, \*\*\* p<0.01



**Figure A.3:** This figure plots the *Event Time* coefficients from separate event studies restricting the sample to all treated schools and all control schools respectively for the outcome of suspension rates. Both sub-sample event studies control for school fixed effects and a vector of school characteristics that includes fraction of Free and Reduced Price Meal (FRPM) students, fraction of underrepresented minority students, and total school enrollment. All lags prior to event time -3 and all leads after event time 3 are dropped from the estimation sample. Standard errors are clustered at the school level.



Figure A.4: This figure shows the event study plots from a Callaway and Sant'Anna (2021) regression where the outcome is suspension rates. From left to right the specifications pictured contain: no demographic controls; a vector of school characteristic controls; and school characteristic controls plus gradespan time trends. In specifications with control, the baseline values of the control variables are to generate inverse propensity-score weights in the aggregation of individual  $2 \times 2$  difference-in-differences ATT estimates up to the event-time level.

	(1)	(2)	(3)
	Baseline	Demographics	Gradespan TT
Treated x ( $\tau = -3$ )	-0.0697	-0.0495	-0.0557
	(0.0836)	(0.0849)	(0.1029)
Treated x ( $\tau = -2$ )	0.0319	0.0813	0.0338
	(0.0820)	(0.0947)	(0.0982)
Treated x ( $\tau = -1$ )	ref.	ref.	ref.
Treated x ( $\tau = 0$ )	-0.0946	-0.0747	-0.0837
	(0.0594)	(0.0574)	(0.0733)
Treated x ( $\tau = 1$ )	-0.176**	-0.187**	-0.201*
	(0.075)	(0.088)	(0.109)
Treated x ( $\tau = 2$ )	-0.218**	-0.137	-0.132
	(0.103)	(0.111)	(0.141)
Treated x ( $\tau = 3$ )	$-0.246^{*}$	-0.237*	-0.207
	(0.137)	(0.136)	(0.186)
F-Stat/Chi-Stat	5.757	5.078	3.519
p-value	0.218	0.279	0.475
Pre-Period Control Mean	1.558	1.558	1.552
Observations	796	796	761

Table A.6: Event Study: Suspensions per Suspended Student

Standard errors in parentheses. Observations are at the school level.

F-stat and p-value come from a test that the coefficients on

Treatment X Event-Time for all post-event years are jointly equal to 0.

\* p<0.1, \*\* p<0.05, \*\*\* p<0.01

	(1)	(2)	(3)
	Baseline	Demographics	Gradespan $\mathrm{TT}$
Treated X Post	-0.1178*	-0.0973	-0.0967
	(0.0636)	(0.0621)	(0.0784)
Pre-Period Control Mean	1.5576	1.5576	1.5518
Observations	796	796	761

Table A.7: Supensions per Suspended Student: Difference-in-Differences

Standard errors in parentheses

Observations are at the school level.

\* p<0.1, \*\* p<0.05, \*\*\* p<0.01

	(1)	(2)	(3)
	All	Non-Defiance	Defiance
Treated X Post	-0.0973	-0.1087	-0.1280
	(0.0621)	(0.1034)	(0.0887)
Pre-Period Control Mean	1.5576	1.7269	1.3616
Observations	796	794	639

Table A.8: Supensions per Suspended Student: Difference-in-Differences by Offense

Observations are at the school level.

\* p<0.1, \*\* p<0.05, \*\*\* p<0.01



**Figure A.5:** This figure shows the event study plots from standard two-way fixed-effects regressions where the outcome is the number of suspension per student for the overall student population of a school. From **left to right** the specifications pictured contain: no demographic controls; a vector of school characteristic controls; and school characteristic controls plus gradespan time trends.

	Standard TWFE			Callaway & Sant'Anna			
	(1)	(2)	(3)	(4)	(5)	(6)	
Treated x ( $\tau = -3$ )	0.0003	0.0003	-0.0011	-0.0077	-0.0055	-0.0072	
	(0.0201)	(0.0201)	(0.0201)	(0.0190)	(0.0199)	(0.0239)	
Treated x ( $\tau = -2$ )	0.0080	0.0080	0.0073	-0.0033	-0.0052	-0.0058	
	(0.0194)	(0.0194)	(0.0201)	(0.0192)	(0.0205)	(0.0242)	
Treated x ( $\tau = -1$ )	ref.	ref.	ref.	ref.	ref.	ref.	
Treated x ( $\tau = 0$ )	-0.0189	-0.0189	-0.0192	-0.0260**	-0.0176	-0.0201	
	(0.0188)	(0.0188)	(0.0187)	(0.0132)	(0.0128)	(0.0160)	
Treated x ( $\tau = 1$ )	-0.045*	-0.045*	-0.045*	-0.040*	-0.029	-0.033	
	(0.025)	(0.025)	(0.024)	(0.021)	(0.021)	(0.026)	
Treated x ( $\tau = 2$ )	-0.045*	-0.045*	-0.046*	-0.048*	-0.035	-0.037	
	(0.027)	(0.027)	(0.026)	(0.027)	(0.027)	(0.034)	
Treated x ( $\tau = 3$ )	-0.040*	-0.040*	-0.041**	-0.052*	-0.047	-0.068*	
	(0.021)	(0.021)	(0.020)	(0.029)	(0.031)	(0.038)	
Baseline Treatment Effect	0.219	0.219	-4.554				
	(0.179)	(0.179)	(7.132)				
School Characteristics	Х	Х	Х				
Gradespan Time Trends			Х				
F-Stat/Chi-Stat				4.376	2.434	3.965	
p-value	0.333	0.333	0.282	0.358	0.656	0.411	
Pre-Period Control Mean	0.069	0.069	0.069	0.069	0.069	0.069	
$R^2$	0.761	0.761	0.768				
Observations	505	505	505	505	505	453	

Table A.9: Event Study: Suspensions per Suspended Student

Standard errors in parentheses. Observations are at the school level. F-stat and p-value come from a test that the coefficients on Treatment X Event-Time for all post-event years are jointly equal to 0. \* p<0.1, \*\* p<0.05, \*\*\* p<0.01

	Sta	andard TW	FE	Callaway & Sant'Anna			
	(1)	(2)	(3)	(4)	(5)	(6)	
Treated X Post	-0.0372**	-0.0379**	-0.0365**	-0.0361**	-0.0284*	-0.0323*	
	(0.0170)	(0.0170)	(0.0160)	(0.0156)	(0.0154)	(0.0193)	
Constant	0.1030***	0.2401	-4.5947				
	(0.0208)	(0.1933)	(7.0437)				
School Characteristics		Х	Х		Х	Х	
Gradespan Time Trends			Х			Х	
Pre-Period Control Mean	0.0689	0.0689	0.0689	0.0689	0.0689	0.0689	
$R^2$	0.761	0.763	0.771				
Observations	505	505	505	505	505	453	

Table A.10: Total Suspensions per Student: Difference-in-Differences

Standard errors in parentheses. Observations are at the school level. \* p<0.1, \*\* p<0.05, \*\*\* p<0.01



**Figure A.6:** This figure show event study plots from standard two-way fixed-effects regressions where the outcome is dropout rates. From **left to right** the specifications pictured contain: no demographic controls; a vector of school characteristic controls; and school characteristic controls plus gradespan time trends.



Figure A.7: This figure plots the  $Treatment \times Event Time$  coefficients from an augmented event study for dropout rates, controlling for matched pair fixed effects and a vector of school characteristics that includes fraction of Free and Reduced Price Meal (FRPM) students, fraction of underrepresented minority students, and total school enrollment. All lags prior to event time -3 and all leads after event time 3 are dropped from the estimation sample. Standard errors are clustered at the school level.

	(1)	(2)	(3)	(4)
	Pooled	Elementary	Middle	High
Treated X Post	-0.0136**	0.0011	-0.0491***	-0.0032
	(0.0059)	(0.0077)	(0.0136)	(0.0045)
Pre-Period Control Mean	0.0649	0.0408	0.0797	0.0715
Observations	867	270	183	407

Table A.11: Suspension Rates: Heterogeneity by Grade-Level Difference-in-Differences

Observations are at the school level.

\* p<0.1, \*\* p<0.05, \*\*\* p<0.01



**Figure A.8:** This figure shows event study plots where the outcome is the school-level dropout rate, for the subsample of high schools. The specifications include a baseline model with no demographic controls (left) and the preferred model which contains a vector of demographic controls (right). Standard errors are clustered at the school level.

	(1)	(2)	(3)
	Pooled	Elementary	Middle
Treated X Post	-0.0012	0.0015	-0.0032**
	(0.0010)	(0.0015)	(0.0013)
Pre-Period Control Mean	0.0084	0.0022	0.0123
Observations	372	113	226

Table A.12: Dropout Rates: Heterogeneity by Grade-Level Difference-in-Differences

Observations are at the school level.

\* p<0.1, \*\* p<0.05, \*\*\* p<0.01



**Figure A.9:** This figure shows event study plots where the outcome is the school-level suspension rate, for the subsample of SBHCs that **report** offering mental health services. The specification pictured contains a vector of school characteristic controls. Standard errors are clustered at the school level.



**Figure A.10:** This figure shows event study plots where the outcome is the school-level dropout rate, for the subsample of SBHCs that **report** offering mental health services. From **left to right** the specifications pictured contain: no demographic controls; a vector of school characteristic controls; and school characteristic controls plus gradespan time trends. Standard errors are clustered at the school level.

	Male Suspension Rate		Female Suspension Rate			
	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline	Demographics	Gradespan $\mathrm{TT}$	Baseline	Demographics	Gradespan $\mathrm{TT}$
Treated x (EY - 3)	-0.0021	0.0000	-0.0071	-0.0018	-0.0007	-0.0047
	(0.0111)	(0.0113)	(0.0132)	(0.0055)	(0.0058)	(0.0076)
Treated x (EY - 2)	-0.0076	-0.0080	-0.0049	-0.0055	-0.0056	-0.0080
	(0.0085)	(0.0087)	(0.0101)	(0.0053)	(0.0055)	(0.0072)
Treated x (EY - 1)						
Treated x Event Year (EY)	-0.013	-0.015	-0.018	-0.003	-0.003	-0.006
	(0.009)	(0.009)	(0.012)	(0.004)	(0.004)	(0.006)
Treated x $(EY + 1)$	-0.017	-0.019*	-0.025*	-0.006	-0.006	-0.011
	(0.010)	(0.011)	(0.014)	(0.006)	(0.006)	(0.008)
Treated x $(EY + 2)$	-0.023	-0.025	-0.030	-0.005	-0.005	-0.009
	(0.015)	(0.017)	(0.021)	(0.006)	(0.006)	(0.008)
Treated x $(EY + 3)$	-0.024	-0.025	-0.029	-0.006	-0.006	-0.001
	(0.016)	(0.019)	(0.025)	(0.007)	(0.007)	(0.012)
F-Stat	2.808	3.199	3.743	1.183	1.017	2.024
p-value	0.422	0.362	0.291	0.757	0.797	0.568
Pre-Period Control Mean	0.087	0.087	0.090	0.041	0.041	0.042
Observations	867	867	812	867	867	812

 Table A.13:
 Suspensions Rates:
 Gender Heterogeneity

Standard errors in parentheses. Observations are at the school level. \* p<0.1, \*\* p<0.05, \*\*\* p<0.01

## **B** Propensity Score Matching and Factor Selection

Recent papers on propensity-score matching have argued for a careful selection of factors in constructing the propensity scores. Smith and Todd (2005) notes that one of the concerns with propensity score matching is that the results may be sensitive to choice of predictors and the specified prediction model. Moreover, the use of "bad predictors" can be equally problematic for matching. In order to avoid the use of bad predictors or a poorly specified model, I focus on evaluating three potential predictors that are grounded in contextual knowledge of SBHCs: the socioeconomic status of students attending a school, the racial composition of the school, and the size of the school. The first two factors are motivated by the credo that first inspired the SBHC model in the 1960s. In their incipience, SBHCs were intended to bridge gaps in healthcare access specifically in low-income areas and for students from racial minority backgrounds (Flaherty and Osher, 2003). Current guidance from the California School-Based Health Alliance suggests that this continues to be a goal of SBHCs in California; specifically, they recommend that SBHCs in California focus on "health care services for children and youth with Medi-Cal coverage" and providing "culturally competent, high-quality, first-contact primary care" with the potential to "reduce health inequities and improve health outcomes for LGBTQ+ youth, low-income youth, and youth of color" (CSBHA, 2022b). While socioeconomic status could be measured through median income, this data is not available at the school level.<sup>36</sup> Instead, I rely on the fraction of students at the school who receive free- or reduced-price lunch, which is a standard proxy for socioeconomic status in the academic literature.<sup>37</sup> The racial composition of a school is measured by the fraction of students from racial minority backgrounds, which includes students who identify as Black, Hispanic or Latinx, Filipino, Pacific Islander, American Indian or Alaskan Native.<sup>38</sup> There is a high correlation between this constructed metric for "racial composition" and the fraction of students receiving free-and reduced price lunch, indicating that this measure is appropriately capturing underlying facets of a school that would increase its likelihood of opening an SBHC.

The third factor, school size, is motivated by background on the process of opening an "onsite" SBHC. Since on-site SBHCs are often located *inside* a physical school building, the ability to construct an SBHC on-site necessitates either a large school or large school campus, both of which are plausibly correlated with a large student body. An additional reason that school size may be a reasonable predictor of opening is that if districts are concerned with improving healthcare access for as many students as possible, they would be incentivized to place the SBHC in a school with a larger concentration of the district's students. To identify the correct set of predictors to use, I run a set of logit models based on equation 3.

<sup>&</sup>lt;sup>36</sup>The most granular geography for Census data on median household income is the census tract level.

 $<sup>^{37}</sup>$ In California, students from households with income below 130% of the federal poverty level qualify for free meals, while students from households that fall between 130% and 185% of the federal poverty level qualify for "reduced-price" meals. (CDE, 2023)

<sup>&</sup>lt;sup>38</sup>This definition is based on definitions of "underrepresented minority" or URM students commonly used at the university level in California.

$$Treated_{s,t} = \alpha_{s,t} + \mathbb{X}_{s,t-1} + \varepsilon_{s,t} \tag{3}$$

where  $Treated_{s,t}$  is an dummy equal to 1 if school s has an active SBHC in year t.  $X_{s,t-1}$  is a vector of lagged predictors for school s in year t-1. In the most saturated specification  $X_{s,t-1}$  contains 1-year lags of the fraction of students qualified for FRPM, the fraction minority students, and the total enrollment. Table B.1 shows the coefficients from a set of logit models. Columns (1) - (3) show logit models for each of the three potential factors separately. The primary takeaway from these first three specifications is that individually, each of these factors is a statistically significant predictor of an increase in the likelihood of a school having an SBHC. The  $\chi^2$  statistics indicate that 1-year lagged enrollment is the most predictive, while fraction of URM students is the least predictive.

Work from Heckman et al. (1998) suggests that one effective method of selecting appropriate predictors is by sequentially adding potential predictors to the model and testing for significance. Following this approach, Table B.1 show three logit models beginning with including only lagged fraction of FRPM students as a predictor in Column (1), and adding lagged enrollment in Column (2) and lagged fraction of minority students in Column (3). Columns (1) and (2) reveal that the lags for both fraction FRPM and total enrollment are statistically significant predictors of opening an SBHC in school s at time t. Column (3) shows that the inclusion of the lagged fraction of minority students provides no additional predictive power. This is not surprising given the high correlation between the fraction of minority students the fraction of FRPM students (with a correlation coefficient of 0.798).

Once the correct predictors have been determined, it remains to select the correct function of these predictors. Table B.2 shows potential linear and non-linear functions of lagged fraction of FRPM students and lagged enrollment. Column (1) shows the original linear function of both variables; Column (2) adds a square term for lagged enrollment to the baseline specification; Column(3) adds a square term for lagged FRPM to the baseline specification; finally, Column (4) includes square terms for both predictors. Column 2 reveals that the addition of a square term for lagged enrollment is statistically significant; however the sign of the coefficient is negative and the  $\chi^2$  statistic for the specification in Column (2) is lower than the  $\chi^2$  statistic for Column (1). This suggests that while the inclusion of a square term for lagged enrollment may be statistically significant, it does not necessarily increase the predictive power of the model.

One further check that is relevant here is which model specification is most predictive *within* each of the three gradespan types. This is worth examining here since the propensity matching process restricts to matching schools in the same gradespan. I focus on the specifications in Columns (1) and (2) of Table B.2, as the two specifications where all predictors are statistically significant. Table B.3 runs each of those logit specifications separately for elementary, middle, and high schools. Columns (1), (3), and (5) reveal that the predictors in the linear specification are consistently significant across all three gradespans. For the non-linear model, all predictors are significant for

	(1)	(2)	(3)
1Y FRPM Lag	2.070***		2.241
	(0.312)		(1.435)
1Y Frac URM Lag	-0.0765		
	(0.363)		
1Y Enrollment Lag	$0.00104^{***}$		0.000828**
	(0.0000816)		(0.000372)
$1Y \ FRPM \ Lag \ Q2$		0.655***	
		(0.202)	
1Y FRPM Lag Q3		0.729***	
		(0.200)	
1Y FRPM Lag Q4		1.073***	
		(0.191)	
1Y Enr Lag Q2		0.390**	
		(0.191)	
1Y Enr Lag Q3		-0.231	
		(0.219)	
1Y Enr Lag Q4		1.073***	
		(0.172)	
$(1YFRPMLag)^2$		( )	-0.172
( 5)			(1.175)
$(1YEnrollmentLag)^2$			7.25e-08
( 5)			(0.000000125)
Constant	-5.935***	-4.793***	-5.925***
	(0.226)	(0.205)	(0.425)
Chi Squared	309.1	162.6	325.3
Observations	11522	11523	11523

 Table B.1: Predicted Likelihood of Having an SBHC (Pooled)

Observations are at the school level.

Covariates are one-year lags relative to a specific cohort event.

\* p<0.1, \*\* p<0.05, \*\*\* p<0.01

	(1)	(2)	(3)	(4)
1Y FRPM Lag	$2.019^{***}$	$2.028^{***}$	2.212	2.241
	(0.252)	(0.252)	(1.444)	(1.435)
1Y Enrollment Lag	$0.00104^{***}$	$0.000830^{**}$	$0.00104^{***}$	$0.000828^{**}$
	(0.0000782)	(0.000374)	(0.0000793)	(0.000372)
$(1YEnrollmentLag)^2$		7.21e-08		7.25e-08
		(0.000000125)		(0.00000125)
$(1YFRPMLag)^2$			-0.156	-0.172
			(1.180)	(1.175)
Constant	-5.956***	-5.873***	-6.004***	-5.925***
	(0.196)	(0.250)	(0.382)	(0.425)
Chi Squared	303.6	314.6	312.5	325.3
Observations	11523	11523	11523	11523

 Table B.2: Predicted Likelihood of Having an SBHC - Model Selection

Observations are at the school level.

Covariates are one-year lags relative to a specific cohort event.

\* p<0.1, \*\* p<0.05, \*\*\* p<0.01

the subsample of elementary schools, but not for middle and high schools. More concerningly, for the sample of middle schools, the addition of the square term for lagged enrollment decreases the  $\chi^2$  statistic, indicating that this model may be less predictive.

In order to ensure the use of a model with consistent predictive power, both for the whole sample and each gradespan subsample, the final logit regression follows equation 3 where the vector  $X_{s,t-1}$  contains the lagged fraction of students qualified for FRPM and the lagged total enrollment of a school. The final matching occurs within gradespan and limits the sample of potential control school districts to those that already have an SBHC; therefore the propensity matching implicitly takes into account grade-levels and underlying openeness to having an SBHC in addition to the selected observable predictors.

	Elementary		Iiddle I		High School	
	(1)	(2)	(3)	(4)	(5)	(6)
1Y FRPM Lag	5.672***	5.723***	4.184***	4.138***	0.525	4.138***
	(0.724)	(0.755)	(1.028)	(0.981)	(0.326)	(0.981)
1Y Enrollment Lag	0.000471	$0.0107^{***}$	-0.00124**	-0.00382***	$0.000610^{***}$	-0.00382***
	(0.000413)	(0.00308)	(0.000521)	(0.00100)	(0.0000830)	(0.00100)
$(1YEnrollmentLag)^2$		-0.00000911***		$0.00000186^{***}$		$0.00000186^{***}$
		(0.00000290)		(0.00000546)		(0.00000546)
Constant	-9.260***	-11.88***	-5.549***	-4.803***	-3.885***	-4.803***
	(0.591)	(1.150)	(0.876)	(0.874)	(0.237)	(0.874)
Chi Squared	86.52	83.77	19.05	28.19	61.44	28.19
Observations	7009	7009	1493	1493	2427	1493

Table B.3: Predicted Likelihood of Having an SBHC (Logit)

Observations are at the school level.

Covariates are one-year lags relative to a specific cohort event.

\* p<0.1, \*\* p<0.05, \*\*\* p<0.01

## C Alternative Control Groups

The primary results in this paper must be cautiously interpreted with respect to the selected control group. As such, there are two pivotal choices in this paper that deserve further justification: (1) the choice of a propensity-matching specification instead of a simple two-way fixed effects model using never or not-yet treated units; and (2) the selection of control schools from the pool of *untreated schools* in districts that have at least one open SBHC.

The use of a propensity-score matching approach to select control schools is motivated by the expectation that there is *selection into treatment*. Specifically, since schools do not randomly receive an SBHC, the schools/districts that choose to open an SBHC may be meaningfully different from those districts that never open an SBHC. Propensity-score matching addresses this concern by matching schools on characteristics that are predictive of the likelihood of opening an SBHC. An alternate method of addressing selection is to use schools that are treated in the *future* as controls for schools having the same underlying propensity to *ever* open an SBHC, the *exact timing* of the open is random. This approach would fail to produce well-matched treatment and control groups if the specific timing at which an SBHC is non-random, and specifically if the timing is driven by district or school-level trends that are correlated with the outcomes of interest. The existence of a pre-trend in suspension rates that is in the same direction as the treatment further complicates the use of future treated schools, since any "decreases" in suspension rates in the control group may be capturing the pre-trend for the future-treated school, and therefore would not be a proper counterfactual for the trajectory of the treated school in absence of treatment.

Figure C.1 shows the separate event studies for treated and control schools using a "futuretreated" control group. Specifically, for the cohort of schools that open an SBHC in year y the control group includes all schools that open an SBHC in year z > y + 3. The imposition of a three year buffer allows for the examination of a three-year post-event window in which all of the "control" schools are pure controls.<sup>39</sup> Figure C.1 shows a visible difference in pre-trends for treated and control schools in this sample, which is in line with the theory that the exact timing of SBHC-openings may not be random.



#### Control: Future-Treated Schools

Note: Shaded area represents the 95% confidence intervals

**Figure C.1:** This figure plots the *Event Time* coefficients from treated and control group event studies for suspension rates (left) and dropout rates (right) using *future-treated* schools as controls. Both sub-sample event studies control for school fixed effects and a vector of school characteristics that includes fraction of Free and Reduced Price Meal (FRPM) students, fraction of underrepresented minority students, and total school enrollment. All lags prior to event time -3 and all leads after event time 3 are dropped from the estimation sample. Standard errors are clustered at the school level.

Having provided some evidence that propensity-score matching is a more appropriate control group than future-treated schools in this setting, we can consider whether the restrictions placed on the propensity score matching process are appropriate. The primary matching process in this paper uses the one-year lagged fraction of FRPM students and one-year lagged school enrollment

<sup>&</sup>lt;sup>39</sup>This model of using future-treated schools that open outside of a certain window is used in several recent papers including Deshpande and Li (2019) and Fadlon and Nielsen (2021). The primary goal of the buffer is to prevent violations of the Stable Unit Treatment Values Assumption (SUTVA) which requires that units do not change their treatment status after the time of the treatment.

as the predictors used to construct the propensity scores; however, it also imposes two additional restrictions that theoretically strengthen the matching: (1) that matches are selected from the same gradespan as the treated school (i.e. a high school with an SBHC can be matched to a high school without an SBHC but not an *middle school*); and (2) that match for a school with an SBHC that opens in year y is selected from the pool of *never-treated* schools in districts that have at least one SBHC that opened in year  $k \le y$ . The first restriction has a natural motivation: both the types of services offered by SBHCs and the outcomes of interest (suspensions aand dropout rates) are likely to differ meaningfully across different gradespans. Therefore gradespan mismatches could be a significant source of bias for my difference-in-differences estimates. The second restriction aims to improve the quality of matches by limiting to districts that have similar "openness" to having an SBHC. An alternate way of accomplishing this it to match within district, following recent recommendations from the propensity-score matching literature to match within the same "local labor market". Figure C.2 shows treated and control group event studies on suspension rates for a sample where controls are selected using *within-district* propensity score matching. The figure reveals that control schools matched from the same district do not have a parallel pre-trend in suspension rates to the set of treated schools.

Finally, we might consider whether the restriction of matching to districts with at least one SBHC is necessary at all. The major concern with selecting concerns from districts that never open an SBHC during the study window is that these districts may be meaningully different on unobservables and policies than districts that ever open an SBHC. In particular, there is a concern that if control schools in these districts have a similar predicted probability of opening an SBHC in year y but opt to not open one, this could be indiction of some alternate policy or program that was implemented in lieu of a school-based health center. If this is the case, parallel pre-trends between the treatment and control groups may be insufficient to satisfy the assumption that the trajectory of outcomes in the control schools represents the correct counterfactual for the expected trajectory of outcomes in treated schools in the absence of treatment. Table C.1 compares the 2012 sample means for a set of school and district-level covariates between districts that ever open an SBHC and districts that never open an SBHC. The sample sizes for each mean are in brackets. This table reveals that ever-treated districts are significantly different from never-treated districts across all covariates. Specifically, ever-treated district tend to be larger on average, with over double the number of high schools, over three times the number of middle schools, and nearly five times the number of elementary schools. While the average school-size is similar, the average school in a treated district has 11 percentage points more students qualifying for Free or Reduced-Price Meals, 16 percentage points more minority students and a lower zip-code-level median household income level (of around \$1,560). These differences on observable characteristics raise further concerns about the number of unobservable characteristics on which these two types of districts could differ.

Figure C.3 shows the treated and control group event studies for a sample constructed through propensity score matching within gradespan in never-treated districts. Here we see that the parallel pre-trend assumption does seem to be met, and that the treatment effect is close to zero with this



#### Control: Matched Schools in Same District

Note: Shaded area represents the 95% confidence intervals

Figure C.2: This figure plots the *Event Time* coefficients from treated and control group event studies for suspension rates (left) and dropout rates (right) using control schools that are selected through 1:1 propensity-score matching from *the same school district* as each treated school. Both sub-sample event studies control for school fixed effects and a vector of school characteristics that includes fraction of Free and Reduced Price Meal (FRPM) students, fraction of underrepresented minority students, and total school enrollment. All lags prior to event time -3 and all leads after event time 3 are dropped from the estimation sample. Standard errors are clustered at the school level.

	Never-Treated Districts	Ever-Treated Districts	p-value
School-Level Covariates			
Fraction FRPM	0.55	0.66	0.000
	[6682]	[1616]	
Fraction Minority	0.57	0.73	0.000
	[6606]	[1620]	
School Size (Total Enrollment)	610.27	621.16	0.467
	[6606]	[1622]	
Zip-Code Level Median Income	29638.70	28073.86	0.000
	[6111]	[1496]	
District-Level Covariates			
Number of Schools	17.49	79.64	0.000
	[6759]	[1629]	
Number of Elementary Schools	11.38	50.38	0.000
	[6095]	[1560]	
Number of Middle Schools	3.20	10.79	0.000
	[5053]	[1468]	
Number of High Schools	5.04	12.65	0.000
	[2307]	[515]	

## Table C.1: Summary Statistics: District & School Characteristics (2012 Data)

*p*-values are from a t-test that the treated and un-treated means are equal.

The number of observations is listed in brackets under each mean.

This comparison omits Los Angeles Unified which had 919 schools in 2012 and is an outlier relative to other treated districts.
control group; however due to the observed dissimilarity between ever-treated and never-treated districts it is difficult to rule out that this zero effect is driven by changes the control schools around the time of the SBHC opening.



Control: Matched Schools in Never-Treated Districts

Note: Shaded area represents the 95% confidence intervals

Figure C.3: This figure plots the *Event Time* coefficients from treated and control group event studies for suspension rates (left) and dropout rates (right) using control schools that are selected through 1:1 propensity-score matching from *never-treated districts* as each treated school. Both sub-sample event studies control for school fixed effects and a vector of school characteristics that includes fraction of Free and Reduced Price Meal (FRPM) students, fraction of underrepresented minority students, and total school enrollment. All lags prior to event time -3 and all leads after event time 3 are dropped from the estimation sample. Standard errors are clustered at the school level.

# D Two-Way Fixed Effects Regressions

	Standard TWFE			Callay	away & Sant'Anna		
	(1)	(2)	(3)	(4)	(5)	(6)	
Treated x ( $\tau = -3$ )	-0.0053	-0.0047	-0.0050	-0.0017	-0.0001	-0.0056	
	(0.0093)	(0.0090)	(0.0091)	(0.0080)	(0.0082)	(0.0097)	
Treated x ( $\tau = -2$ )	-0.0093	-0.0096	-0.0096	-0.0063	-0.0066	-0.0062	
	(0.0068)	(0.0070)	(0.0069)	(0.0060)	(0.0061)	(0.0076)	
Treated x ( $\tau = -1$ )	ref.	ref.	ref.	ref.	ref.	ref.	
Treated x ( $\tau = 0$ )	-0.0092	-0.0084	-0.0079	-0.0085	-0.0091*	$-0.0119^{*}$	
	(0.0070)	(0.0071)	(0.0072)	(0.0053)	(0.0054)	(0.0071)	
Treated x ( $\tau = 1$ )	-0.017**	-0.017**	-0.016**	-0.012	-0.012	-0.018*	
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.010)	
Treated x ( $\tau = 2$ )	-0.020**	-0.019*	-0.019*	-0.014	-0.015	-0.019	
	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.014)	
Treated x ( $\tau = 3$ )	-0.017**	-0.019**	-0.019**	-0.015	-0.016	-0.015	
	(0.008)	(0.009)	(0.009)	(0.010)	(0.011)	(0.015)	
Baseline Treatment Effect	0.233***	$0.194^{**}$	3.332				
	(0.008)	(0.086)	(2.512)				
School Characteristics		Х	Х		Х	Х	
Gradespan Time Trends			Х			Х	
F-Stat/Chi-Stat	1.332	1.417	1.340	3.090	3.161	4.123	
p-value	0.261	0.231	0.258	0.543	0.531	0.390	
Pre-Period Control Mean	0.065	0.065	0.065	0.065		0.067	
$R^2$	0.807	0.811	0.812				
Observations	867	867	867	867	867	812	

Table D.1: Event Study: Suspensions Rates

Standard errors in parentheses. Observations are at the school level.

F-stat and p-value come from a test that the coefficients on

Treatment X Event-Time for all post-event years are jointly equal to 0.

	Star	ndard TW	FE	Callaway & Sant'Anna			
	(1)	(2)	(3)	(4)	(5)	(6)	
Treated X Post	-0.0098	-0.0098	-0.0098	-0.0130**	-0.0136**	-0.0171**	
	(0.0067)	(0.0069)	(0.0070)	(0.0057)	(0.0059)	(0.0074)	
Constant	$0.0700^{***}$	0.0159	0.6814				
	(0.0076)	(0.0588)	(0.6324)				
School Characteristics		Х	Х		Х	Х	
Gradespan Time Trends			Х			Х	
Pre-Period Control Mean	0.0649	0.0649	0.0649	0.0649	0.0649	0.0668	
Observations	854	854	854	867	867	812	

 Table D.2:
 Suspension Rates:
 Difference-in-Differences

	Sta	andard TW	FE	Callav	vay & Sant'.	Anna
	(1)	(2)	(3)	(4)	(5)	(6)
Treated x ( $\tau = -3$ )	0.0006	0.0008	0.0007	0.0006	0.0007	0.0012
	(0.0030)	(0.0031)	(0.0031)	(0.0025)	(0.0025)	(0.0026)
Treated x ( $\tau = -2$ )	0.0012	0.0015	0.0014	0.0018	0.0021	0.0024
	(0.0034)	(0.0034)	(0.0034)	(0.0028)	(0.0028)	(0.0028)
Treated x ( $\tau = -1$ )	ref.	ref.	ref.	ref.	ref.	ref.
Treated x ( $\tau = 0$ )	-0.0050	-0.0049	-0.0050	-0.0011	0.0009	0.0001
	(0.0036)	(0.0039)	(0.0039)	(0.0016)	(0.0023)	(0.0020)
Treated x ( $\tau = 1$ )	-0.002	-0.002	-0.002	-0.001	0.000	-0.001
	(0.003)	(0.003)	(0.003)	(0.001)	(0.002)	(0.001)
Treated x ( $\tau = 2$ )	-0.004	-0.004	-0.004	-0.004	-0.003	-0.005
	(0.006)	(0.006)	(0.006)	(0.003)	(0.002)	(0.004)
Treated x ( $\tau = 3$ )	-0.018***	-0.018***	-0.019***	-0.016***	-0.012***	
	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	
Baseline Treatment Effect	$0.026^{***}$	$0.080^{*}$	0.146			
	(0.003)	(0.048)	(0.902)			
School Characteristics		Х	Х		Х	Х
Gradespan Time Trends			Х			Х
F-Stat/Chi-Stat	26.296	9.958	7.146	37.173	24.087	2.522
p-value	0.000	0.000	0.000	0.000	0.000	0.471
Pre-Period Control Mean	0.008	0.008	0.008	0.008	0.008	0.008
$R^2$	0.723	0.727	0.727			
Observations	383	383	383	372	372	351

 Table D.3: Dropout Rates Event Study Specifications

F-stat and p-value come from a test that the coefficients on

Treatment X Event-Time for all post-event years are jointly equal to 0.

	Star	ndard TW	FE	Callaw	Callaway & Sant'Anna		
	(1)	(2)	(3)	(4)	(5)	(6)	
Treated X Post	-0.0034	-0.0035	-0.0038	-0.0020**	-0.0012	-0.0016	
	(0.0026)	(0.0026)	(0.0027)	(0.0010)	(0.0010)	(0.0010)	
Constant	0.0131***	-0.0000	0.4464				
	(0.0034)	(0.0110)	(0.2847)				
School Characteristics		Х	Х		Х	Х	
Gradespan Time Trends			Х			Х	
Pre-Period Control Mean	0.0083	0.0083	0.0083	0.0084	0.0084	0.0082	
$R^2$	0.720	0.722	0.725				
Observations	378	378	378	372	372	351	

 Table D.4:
 Dropout Rates Difference-in-Differences Specifications

	(	Standard T	WFE	Cal	laway & Sa	ant'Anna
	(1)	(2)	(3)	(4)	(5)	(6)
	All	Defiance	Non-Defiance	All	Defiance	Non-Defiance
Treated x ( $\tau = -3$ )	-0.0047	0.0034	-0.0082	-0.0017	0.0010	-0.0038
	(0.0090)	(0.0054)	(0.0074)	(0.0080)	(0.0060)	(0.0061)
Treated x ( $\tau = -2$ )	-0.0096	-0.0055	-0.0041	-0.0063	-0.0057	-0.0014
	(0.0070)	(0.0051)	(0.0057)	(0.0060)	(0.0045)	(0.0049)
Treated x ( $\tau = -1$ )	ref.	ref.	ref.	ref.	ref.	ref.
Treated x ( $\tau = 0$ )	-0.0084	-0.0018	-0.0067	-0.0085	-0.0028	-0.0068
	(0.0071)	(0.0047)	(0.0074)	(0.0053)	(0.0051)	(0.0055)
Treated x ( $\tau = 1$ )	-0.017**	-0.007	-0.010	-0.012	-0.004	-0.009
	(0.008)	(0.004)	(0.009)	(0.008)	(0.004)	(0.009)
Treated x ( $\tau = 2$ )	-0.019*	-0.004	-0.015	-0.014	0.001	-0.013
	(0.010)	(0.005)	(0.010)	(0.010)	(0.005)	(0.011)
Treated x ( $\tau = 3$ )	-0.019**	0.001	-0.020*	-0.015	0.005	-0.013
	(0.009)	(0.007)	(0.010)	(0.010)	(0.008)	(0.013)
Baseline Treatment Effect	$0.194^{**}$	0.053	0.142			
	(0.086)	(0.069)	(0.122)			
School Characteristics	Х	Х	Х	Х	Х	Х
F-Stat/Chi-Stat	1.417	1.078	1.145	3.090	2.098	3.004
p-value	0.231	0.370	0.338	0.543	0.718	0.557
Pre-Period Control Mean	0.023	0.042	0.023	0.065	0.044	0.023
$R^2$	0.811	0.779	0.580			
Observations	867	867	867	867	812	867

Table D.5: Event Study: Suspensions by Offense-Types

F-stat and p-value come from a test that the coefficients on

Treatment X Event-Time for all post-event years are jointly equal to 0.

	(1)	(2)	(3)
	Any Offense	Defiance	Non-Defiance
Treated X Post	-0.0098	-0.0081	-0.0017
	(0.0069)	(0.0078)	(0.0039)
Fraction FRPM	0.0134	0.0154	-0.0018
	(0.0347)	(0.0324)	(0.0157)
Fraction Minority	0.0880	-0.0451	0.1331***
	(0.0629)	(0.0814)	(0.0502)
School Size	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)
Constant	0.016	0.083	-0.067**
	(0.059)	(0.066)	(0.030)
School Characteristics	Yes	Yes	Yes
Pre-Period Control Mean	0.0649	0.0230	0.0420
$R^2$	0.807	0.573	0.782
Observations	854	854	854

 Table D.6:
 Suspensions Rates: Heterogeneity by Offense Type (DiD)

	Sta	ndard TW	FE	Callav	Callaway & Sant'Anna		
	(1)	(2)	(3)	(4)	(5)	(6)	
Treated x ( $\tau = -3$ )	-0.0931	-0.0310	-0.0641	-0.0697	-0.0495	-0.0557	
	(0.1100)	(0.0941)	(0.0929)	(0.0836)	(0.0849)	(0.1029)	
Treated x ( $\tau = -2$ )	0.0260	0.0711	0.0646	0.0319	0.0813	0.0338	
	(0.1075)	(0.0961)	(0.0890)	(0.0820)	(0.0947)	(0.0982)	
Treated x ( $\tau = -1$ )	ref.	ref.	ref.	ref.	ref.	ref.	
Treated x ( $\tau = 0$ )	-0.2065	-0.1367	-0.1223	-0.0946	-0.0747	-0.0837	
	(0.1399)	(0.1026)	(0.0986)	(0.0594)	(0.0574)	(0.0733)	
Treated x ( $\tau = 1$ )	-0.152	-0.103	-0.073	-0.176**	-0.187**	-0.201*	
	(0.123)	(0.095)	(0.093)	(0.075)	(0.088)	(0.109)	
Treated x ( $\tau = 2$ )	-0.210	-0.147	-0.130	-0.218**	-0.137	-0.132	
	(0.134)	(0.118)	(0.113)	(0.103)	(0.111)	(0.141)	
Treated x ( $\tau = 3$ )	-0.232	-0.177	-0.166	-0.246*	-0.237*	-0.207	
	(0.159)	(0.129)	(0.126)	(0.137)	(0.136)	(0.186)	
Baseline Treatment Effect	5.881***	-0.509	17.366				
	(0.234)	(2.226)	(20.651)				
School Characteristics		Х	Х		Х	Х	
Gradespan Time Trends			Х			Х	
F-Stat/Chi-Stat	0.771	0.696	0.697	5.757	5.078	3.519	
p-value	0.546	0.596	0.595	0.218	0.279	0.475	
Pre-Period Control Mean	1.558	1.558	1.558	1.558	1.558	1.552	
$R^2$	0.545	0.598	0.631				
Observations	808	808	808	796	796	761	

 Table D.7:
 Event Study: Suspensions per Suspended Student

F-stat and p-value come from a test that the coefficients on

Treatment X Event-Time for all post-event years are jointly equal to 0.

	Sta	ndard TWF	È	Callaway & Sant'Anna			
	(1)	(2)	(3)	(4)	(5)	(6)	
Treated X Post	-0.1211*	-0.1131	-0.1117	-0.1178*	-0.0973	-0.0967	
	(0.0703)	(0.0693)	(0.0702)	(0.0636)	(0.0621)	(0.0784)	
Constant	1.5906***	2.5159***	4.2826				
	(0.0876)	(0.4495)	(5.0535)				
School Characteristics		Х	Х		Х	Х	
Gradespan Time Trends			Х			Х	
Pre-Period Control Mean	1.5576	1.5576	1.5576	1.5576	1.5576	1.5518	
$\mathbb{R}^2$	0.621	0.627	0.628				
Observations	797	797	797	796	796	761	

 Table D.8:
 Supensions per Suspended Student:
 Difference-in-Differences

# E Full Sample Regressions

	Sta	ndard TW	FE	Callav	Callaway & Sant'Anna		
	(1)	(2)	(3)	(4)	(5)	(6)	
Treated x ( $\tau = -3$ )	-0.0043	-0.0036	-0.0036	-0.0010	-0.0015	-0.0066	
	(0.0083)	(0.0083)	(0.0083)	(0.0073)	(0.0075)	(0.0095)	
Treated x ( $\tau = -2$ )	-0.0063	-0.0054	-0.0055	-0.0047	-0.0043	-0.0049	
	(0.0065)	(0.0064)	(0.0064)	(0.0058)	(0.0060)	(0.0073)	
Treated x ( $\tau = -1$ )	ref.	ref.	ref.	ref.	ref.	ref.	
Treated x ( $\tau = 0$ )	-0.0086	-0.0078	-0.0076	-0.0071	-0.0073	-0.0093	
	(0.0061)	(0.0061)	(0.0061)	(0.0047)	(0.0047)	(0.0060)	
Treated x ( $\tau = 1$ )	-0.016**	-0.015**	-0.015**	-0.009	-0.011	-0.016*	
	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)	(0.009)	
Treated x ( $\tau = 2$ )	-0.017**	-0.017**	-0.016**	-0.011	-0.012	-0.016	
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.011)	
Treated x ( $\tau = 3$ )	-0.016**	-0.015**	-0.015**	-0.012	-0.016*	-0.016	
	(0.007)	(0.007)	(0.006)	(0.008)	(0.009)	(0.010)	
Baseline Treatment Effect	$0.160^{***}$	0.204	-0.309				
	(0.007)	(0.136)	(0.763)				
School Characteristics		Х	Х		Х	Х	
Gradespan Time Trends			Х			Х	
F-Stat/Chi-Stat	1.640	1.503	1.598	3.215	4.624	4.244	
p-value	0.167	0.203	0.177	0.523	0.328	0.374	
Pre-Period Control Mean	0.041	0.041	0.041	0.041		0.042	
$R^2$	0.834	0.835	0.837				
Observations	991	991	991	991	991	909	

 Table E.1:
 Event Study: Suspensions Rates

Standard errors in parentheses. Observations are at the school level.

F-stat and p-value come from a test that the coefficients on

Treatment X Event-Time for all post-event years are jointly equal to 0.

	Sta	ndard TWI	FE	Callaway & Sant'Anna			
	(1)	(2)	(3)	(4)	(5)	(6)	
Treated X Post	-0.0099*	-0.0100*	-0.0099*	-0.0112**	-0.0126**	-0.0155**	
	(0.0056)	(0.0057)	(0.0053)	(0.0048)	(0.0050)	(0.0062)	
Constant	$0.0568^{***}$	$0.0704^{**}$	-0.2965				
	(0.0028)	(0.0320)	(0.2929)				
School Characteristics		Х	Х		Х	Х	
Gradespan Time Trends			Х			Х	
Pre-Period Control Mean	0.0415	0.0415	0.0415	0.0415	0.0415	0.0416	
Observations	977	977	977	991	991	909	

 Table E.2:
 Suspension Rates:
 Difference-in-Differences

	Sta	andard TW	FE	Callaw	Callaway & Sant'Anna		
	(1)	(2)	(3)	(4)	(5)	(6)	
Treated x ( $\tau = -3$ )	0.0007	0.0009	0.0010	0.0007	0.0012	0.0018	
	(0.0031)	(0.0031)	(0.0031)	(0.0024)	(0.0024)	(0.0029)	
Treated x ( $\tau = -2$ )	0.0019	0.0022	0.0022	0.0026	0.0026	0.0026	
	(0.0033)	(0.0033)	(0.0033)	(0.0027)	(0.0027)	(0.0029)	
Treated x ( $\tau = -1$ )	ref.	ref.	ref.	ref.	ref.	ref.	
Treated x ( $\tau = 0$ )	-0.0017	-0.0015	-0.0015	0.0008	-0.0005	0.0013	
	(0.0037)	(0.0039)	(0.0040)	(0.0020)	(0.0022)	(0.0016)	
Treated x ( $\tau = 1$ )	-0.001	-0.000	-0.000	0.001	0.001	0.001	
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	
Treated x ( $\tau = 2$ )	-0.003	-0.002	-0.002	0.001	0.000	-0.001	
	(0.005)	(0.005)	(0.005)	(0.003)	(0.004)	(0.007)	
Treated x ( $\tau = 3$ )	-0.013***	-0.011***	-0.011***	-0.010***	-0.009*		
	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)		
Baseline Treatment Effect	0.043***	0.075	-0.082				
	(0.003)	(0.063)	(0.966)				
School Characteristics		Х	Х		Х	Х	
Gradespan Time Trends			Х			Х	
F-Stat/Chi-Stat	4.562	4.313	3.835	14.193	6.306	0.738	
p-value	0.002	0.003	0.006	0.007	0.177	0.864	
Pre-Period Control Mean	0.012	0.012	0.012	0.012	0.012	0.012	
$R^2$	0.815	0.816	0.817				
Observations	492	492	492	484	484	435	

 Table E.3: Dropout Rates Event Study Specifications

F-stat and p-value come from a test that the coefficients on

Treatment X Event-Time for all post-event years are jointly equal to 0.

	Standard TWFE			Callav	allaway & Sant'Anna		
	(1)	(2)	(3)	(4)	(5)	(6)	
Treated X Post	-0.0019	-0.0018	-0.0020	-0.0006	-0.0009	-0.0002	
	(0.0024)	(0.0025)	(0.0026)	(0.0013)	(0.0014)	(0.0013)	
Constant	$0.0157^{***}$	-0.0091	0.5060				
	(0.0020)	(0.0104)	(0.5462)				
School Characteristics		Х	Х		Х	Х	
Gradespan Time Trends			Х			Х	
Pre-Period Control Mean	0.0120	0.0120	0.0120	0.0120	0.0120	0.0115	
$R^2$	0.816	0.819	0.819				
Observations	486	486	486	484	484	435	

 Table E.4:
 Dropout Rates Difference-in-Differences Specifications

Standard errors in parentheses

Observations are at the school level.

	Standard TWFE			Callaway & Sant'Anna			
	(1)	(2)	(3)	(4)	(5)	(6)	
	All	Non-Defiance	Defiance	All	Non-Defiance	Defiance	
Treated x ( $\tau = -3$ )	-0.0036	-0.0025	-0.0011	-0.0010	-0.0060	0.0004	
	(0.0083)	(0.0049)	(0.0064)	(0.0073)	(0.0062)	(0.0056)	
Treated x ( $\tau = -2$ )	-0.0054	-0.0078*	0.0024	-0.0047	-0.0083**	0.0016	
	(0.0064)	(0.0046)	(0.0045)	(0.0058)	(0.0042)	(0.0048)	
Treated x ( $\tau = -1$ )	ref.	ref.	ref.	ref.	ref.	ref.	
Treated x ( $\tau = 0$ )	-0.0078	0.0007	-0.0085	-0.0071	0.0000	-0.0073	
	(0.0061)	(0.0043)	(0.0064)	(0.0047)	(0.0048)	(0.0046)	
Treated x ( $\tau = 1$ )	-0.015**	-0.004	-0.011	-0.009	-0.002	-0.010	
	(0.007)	(0.004)	(0.008)	(0.006)	(0.004)	(0.008)	
Treated x ( $\tau = 2$ )	-0.017**	-0.002	-0.015*	-0.011	0.004	-0.015*	
	(0.008)	(0.004)	(0.008)	(0.008)	(0.005)	(0.009)	
Treated x ( $\tau = 3$ )	-0.015**	-0.001	-0.014*	-0.012	-0.003	-0.015	
	(0.007)	(0.006)	(0.008)	(0.008)	(0.007)	(0.010)	
Baseline Treatment Effect	0.204	0.067	0.138				
	(0.136)	(0.127)	(0.111)				
School Characteristics	Х	Х	Х	Х	Х	Х	
F-Stat/Chi-Stat	1.503	0.607	1.118	3.215	2.047	4.815	
p-value	0.203	0.658	0.350	0.523	0.727	0.307	
Pre-Period Control Mean	0.041	0.028	0.013	0.041	0.029	0.013	
$R^2$	0.835	0.804	0.643				
Observations	991	991	991	991	909	991	

 Table E.5:
 Event Study: Suspensions by Offense-Types

F-stat and p-value come from a test that the coefficients on

Treatment X Event-Time for all post-event years are jointly equal to 0.

	Standard TWFE			Callaway & Sant'Anna		
	(1)	(2)	(3)	(4)	(5)	(6)
	All	Non-Defiance	Defiance	All	Non-Defiance	Defiance
Treated X Post	-0.0100*	0.0017	-0.0117*	-0.0126**	-0.0027	-0.0099*
	(0.0057)	(0.0033)	(0.0066)	(0.0050)	(0.0030)	(0.0052)
Pre-Period Control Mean	0.0415	0.0284	0.0131	0.0415	0.0284	0.0131
Observations	977	977	977	991	991	991

Table E.6: Suspensions Rates: Heterogeneity by Offense Type (DiD)

	Standard TWFE			Callaway & Sant'Anna		
	(1)	(2)	(3)	(4)	(5)	(6)
Treated x ( $\tau = -3$ )	-0.1304	-0.0648	-0.0657	-0.0458	-0.0388	-0.0398
	(0.1111)	(0.0872)	(0.0856)	(0.0743)	(0.0724)	(0.0866)
Treated x ( $\tau = -2$ )	0.0116	0.0807	0.0783	0.0238	0.0371	0.0082
	(0.1102)	(0.0856)	(0.0820)	(0.0757)	(0.0803)	(0.0808)
Treated x ( $\tau = -1$ )	ref.	ref.	ref.	ref.	ref.	ref.
Treated x ( $\tau = 0$ )	-0.1787	-0.1039	-0.1007	-0.0929*	-0.0799	-0.0773
	(0.1243)	(0.0909)	(0.0882)	(0.0553)	(0.0554)	(0.0682)
Treated x ( $\tau = 1$ )	-0.224**	-0.164*	-0.152*	-0.171**	-0.146**	-0.146*
	(0.112)	(0.083)	(0.081)	(0.068)	(0.069)	(0.084)
Treated x ( $\tau = 2$ )	-0.189	-0.134	-0.114	-0.222**	-0.182*	-0.175
	(0.124)	(0.104)	(0.096)	(0.095)	(0.099)	(0.114)
Treated x ( $\tau = 3$ )	-0.200*	-0.124	-0.140	$-0.216^{*}$	-0.228**	-0.260*
	(0.113)	(0.108)	(0.092)	(0.113)	(0.115)	(0.145)
Baseline Treatment Effect	5.233***	1.418	-2.350			
	(0.215)	(2.165)	(7.237)			
School Characteristics		Х	Х		Х	Х
Gradespan Time Trends			Х			Х
F-Stat/Chi-Stat	1.070	0.973	0.944	6.877	5.174	3.902
p-value	0.373	0.424	0.440	0.143	0.270	0.419
Pre-Period Control Mean	1.366	1.366	1.366	1.369	1.369	1.371
$R^2$	0.547	0.610	0.641			
Observations	853	853	853	835	835	775

 Table E.7:
 Event Study: Suspensions per Suspended Student

Standard errors in parentheses

Standard errors in parentheses. Observations are at the school level.

F-stat and p-value come from a test that the coefficients on

Treatment X Event-Time for all post-event years are jointly equal to 0.

	Standard TWFE			Callaway & Sant'Anna		
	(1)	(2)	(3)	(4)	(5)	(6)
Treated X Post	-0.1294**	-0.1245**	-0.1218**	-0.1151**	-0.0926	-0.0870
	(0.0595)	(0.0589)	(0.0554)	(0.0574)	(0.0595)	(0.0702)
Constant	1.4680***	2.0112***	-1.1904			
	(0.0573)	(0.3057)	(2.5666)			
School Characteristics		Х	Х		Х	Х
Gradespan Time Trends			Х			Х
Pre-Period Control Mean	1.3661	1.3661	1.3661	1.3686	1.3686	1.3708
$R^2$	0.632	0.634	0.638			
Observations	841	841	841	835	835	775

 Table E.8:
 Supensions per Suspended Student:
 Difference-in-Differences

Standard errors in parentheses

Observations are at the school level.

## F California Department of Education Data Descriptions

CDE Data Category	Offense	California Edu. Codes
Violent Incident (Injury)	Sexual Battery/Assault	48915(c)(4), 48900(n)
	Caused Physical Injury	48915(a)(1)(A)
	Committed Assault or Battery on a School Employee	48915(a)(1)(E)
	Used Force or Violence	48900(a)(2)
	Committed an act of Hate Violence	48900.3
	Hazing	48900(q)
Weapons Possession	Possession, Sale, Furnishing a Firearm	48915(c)(1)
	Possession, Sale, Furnishing a Firearm or Knife	48900(b)
	Brandishing a Knife	48915(c)(2)
	Possession of a Knife or Dangerous Object	48915(a)(1)(B)
	Possession of an Explosive	48915(c)(5)
Illicit Drug-Related	Sale of Controlled Substance	48915(c)(3)
	Possession of Controlled Substance	48915(a)(1)(C)
	Possession, Use, Sale, or Furnishing a Controlled Sub-	48900(c)
	stance, Alcohol, Intoxicant	
	Offering, Arranging, or Negotiating Sale of Controlled Substances, Alcohol, Intoxicants	48900(d)
	Offering, Arranging, or Negotiating Sale of Drug Para- phernalia	48900(j)
	Offering, Arranging, or Negotiating Sale of Soma	48900(p)
Other Reasons	Possession of an Imitation Firearm	48900(m)
	Possession or Use of Tobacco Products	48900(h)
	Property Damage	48900(f)
	Robbery or Extortion	48915(a)(1)(D)
	Property Theft	48900(g)
	Received Stolen Property	48900(l)
Defiance-Only	Disruption, Defiance	48900(k)(1)

#### F.1 Suspension Offense Categories

**Table F.1:** This table shows the various offenses that are included in each "category" of suspensions defined by the California Department of Education. The third column shows the corresponding codes from *California Education Code* §48900 - 48927. The original data definitions can be found at: https://www.cde.ca.gov/ds/ad/fssd.asp

### F.2 California Healthy Kids Survey

The core module of the California Healthy Kids Survey (CHKS) consists of around 155 questions that are selected to assess three pillars of developmental supports that research has linked to "positive academic, psychosocial, and health outcomes among youth, even in high-risk environments": positive academic relationships; high expectations (academic and behavioral); and opportunities for meaningful participation and decision-making.<sup>40</sup>. Several papers have attempted to validate the psychometric properties of subsets of CHKS questions. One such paper comes from researchers at WestEd, the organization that lead the development of the CHKS (Mahecha and Hanson, 2020). This paper proposes the construction of a set of nine indices as weighted averages of subsets of the CHKS questions and verifies the internal consistency reliability and item bias of the constructs. I focus on four indices that are most likely to be correlated with mental health: caring staff-student relationships, school connectedness, delinquency, and substance use at school. Table F.2 partially reproduces a table from Mahecha and Hanson (2020) that lists the questions included in each index and the weight assigned to each question.<sup>41</sup>

In order to construct an index for each construct that takes on the same values as the questions with the index, I scale all weights to sum to one prior to taking a weighted average across all question in the index. Equation F.2 shows the equation for a given index, c, as a weighted average of a set of questions  $\{Q_{c,i}\}$ 

$$I_c = \sum_{\forall Q_{c,i}} (Q_{c,i}) \frac{\omega_{c,i}}{\sum_{\forall c,j} \omega_{c,j}}$$

 $\frac{\omega_i}{\sum_{\forall j} \omega_j}$  represents the scaled weight and  $Q_{c,i}$  is the value for question *i* of construct *c*.

Responses that are missing answers for all questions in an index are assigned an index value of *missing*. For cases where an individual response contains missing answers for some but not all questions in an index, the index is re-scaled by dividing the value of the index by the sum of the weights on the questions with non-missing responses. This amounts to rescaling the weights on the questions that are answered to add up to one. I verify that this rescaling does not bias the index values by comparing the rescaled indices to indices constructed for the subset of responses with no missing questions.

Finally, for my two measures of mental health, I use two questions that are similar to the types of questions commonly used in other surveys measuring mental health. In particular, discussions with researchers who have worked closely with the CHKS suggest that the two questions on the CHKS that directly ask about mental health are drawn from similar surveys such as the Youth Risk Behavior Surveillance System (YRBSS). For these two questions, I use the individual question values as there is no obvious precedent for the conversion of these questions into a weighted index.

<sup>&</sup>lt;sup>40</sup>Source: https://calschls.org/about/the-surveys/

<sup>&</sup>lt;sup>41</sup>In Mahecha and Hanson (2020) each question is assigned a standardized factor loading from a confirmatory factor analysis model. As is standard in CFA models, the factor loading for each question comes from the correlation between that question and the underlying construct being measured. In constructing an index from a set of questions, each question is weighted by the factor loading to account for differences in how well each question captures the underlying construct.

California Healthy Kids Survey Item	Weight			
Caring Staff-Student Relationships				
teacher or adult who really cares about me	0.806			
teacher or adult who tells me when I do a good job	0.836			
teacher or adult who notices when I'm not there	0.737			
teacher or adult who always wants me to do my best	0.864			
teacher or adult who listens to me when I have something to say				
teacher or adult who believes that I will be a successful student	0.873			
Caring Staff-Student Relationships				
I feel close to people at this school	0.649			
I am happy to be at this school	0.835			
I feel like I am part of this school	0.855			
The teachers at this school treat students fairly	0.710			
I feel safe in my school	0.735			
Delinquency				
been in a physical fight at school (12 months)				
been offered, sold, or given an illegal drug at school (12 months)				
damaged school property on purpose at school (12 months)				
carried a gun at school (12 months)				
carried any other weapon at school (12 months)				
been threatened or injured with a weapon at school (12 months)	0.870			
seen someone carrying a gun, knife, or other weapon at school (12 months)	0.720			
been threatened with harm or injury at school (12 months)	0.885			
Substance Use at School				
cigarettes on school property (30 days)	0.939			
smokeless tobacco on school property $(30 \text{ days})$				
electronic cigarettes, e-cigarette on school property (30 days)				
at least one drink of alcohol on school property (30 days)	0.874			
marijuana on school property (30 days)	0.910			
any other drug, pill, or medicine to get "high" on school property (30 days)	0.936			

**Table F.2:** This table lists the "item" and associated weight for each of the four indices constructed to measure school climate and socioemotional well-being. The table structure and contents are a reproduction of the table on pages 38-40 of Mahecha and Hanson (2020).