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

Anjali Pai

November 23, 2023

Link to current version of paper

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 school-based 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 1.1 percentage points (27% of the baseline suspension rate) within 3 years of the opening when compared to matched schools. A heterogeneity analysis reveals 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.[1]

[1]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, and seminar participants at UC San Diego.

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). 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 suggests that there may be barriers to take-up of existing mental health services. The existence of such barriers is especially important given recent state and federal initiatives to increase funding and resources for adolescent mental health services. For example, in 2022, California Governor Gavin Newsom proposed an initiative that would allocate 4.7 billion dollars to improving statewide mental health systems for adolescents.[2] Efficient allocation of funds requires an understanding of what types of programs are the most effective at reaching adolescents and addressing mental-health issues.

A 2019 report from the National Center for Behavioral Health reports that the three primary barriers to take-up of mental health services are 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 barriers: school-based health centers. School-based health centers (known in shorthand as SBHCs) 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.[3]

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. This paper attempts to isolate the impact of school-based health centers on mental health by examining the effect of school-based

[2]Source: *California Awards $30.5 Million for Kids' Mental Health, and Support for Parents and Family Caregivers*, Office of Governor Gavin Newsom.

[3]Source: These statistics come from the 2016-2017 National Census of School-Based Health Centers (Love et al., 2019).

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-in-differences 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 between 0.9-1.1 percentage points, with my preferred regression specification estimating a 1.1 percentage point decrease. This is a large magnitude effect, amounting to a 27% decrease from the control group baseline rate. 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 significantly bias my primary estimates.

Exploring heterogeneity in 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.[4] 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

[4]The California Department of Education defines six total categories of offenses that may lead to a suspension: (1) Violent Incident (Injury); (2) Violent Incident (No Injury); (3) Weapons Possession; (4) Illicit Drug-Related; (5) 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 24 lists the offenses in each category.

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 cross-school 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 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 their 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.

Komisarow and Hemelt find 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 Komisarow and Hemelt study a very different intervention (telemedicine as opposed to in-person services), their 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 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; and finally Section 9 discusses potential policy implications and proposes avenues for future research.

2 Background

The setting for this analysis is the state of California, which serves over 5.8 million K-12 students.[5] According to a 2021 report from the UCLA Health Center for Policy Research, the mental health trends for adolescents in California have closely mirrored national trends in the past decade. Nearly 45% of youths aged 12-17 in California reported having struggled with mental health issues in 2021. Despite a statewide push for improved mental health systems, California is one of the few states that does not currently provide state funding for School-Based Health Centers, suggesting that the large costs of these centers may be a deterrent to state investment (Love et al., 2019). As of

[5]Source: *About School-Based Health Centers*, California School Based Health Alliance

the 2022-2023 school year, there were 346 active School-Based Health Centers in California.[6] The California School-Based Health Alliance (CSBHA) defines school-based health centers as "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, 2022). Health services in schools that do not have an SBHC is most commonly provided by registered 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.[7]

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, 2022). 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, 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 ($\sim 2\%$) (vis, 2022).

The unique features of the school-based health centers make them well-equipped to increase take-up 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 SBHC-model 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

[6]Source: *Fingertip Facts on Education in California*, California Department of Education

[7]Source: *School-Based Health & Wellness Centers in California: A Growing Trend*, California School Based Health Alliance

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.11 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 one-on-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.[8]

Beyond just targeting the expected barriers to take-up of mental health services, the adolescent-focused 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 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

[8] These case studies come from a report compiled by Lisa Eisenberg, formerly of the California School Based Health Alliance.

from the California Healthy Kids Survey (CHKS).[9] 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).[10]

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)).[11] 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 paper limit to the set of openings between 2012-2019 (indicated by the dark purple region in Figure 1). The available survey data from the CHKS extends back to 1998 (the pink shaded region), however when used in combination with suspension and dropouts data, I am once again forced to limit my sample to the later years.

3.2 Constructing the Study Sample

Constructing the analysis dataset involved a few key steps that I will discuss in this section. The first step was to merge the data on SBHC openings to the CDE data on school-level suspensions, expulsions, dropout rates. The first necessary involved matching each SBHC to a "principal school". For on-site SBHCs, the "principal school" is the school in which the SBHC is physically located. For off-site and mobile vans, I consider the "principal school" to be the school in closest geographic proximity to the SBHC. The SBHC dataset contains text fields with the name of the "school served"

[9]The California School-Based Health Alliance is a state affiliate of the National School Based Health Alliance.

[10]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.

[11]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.

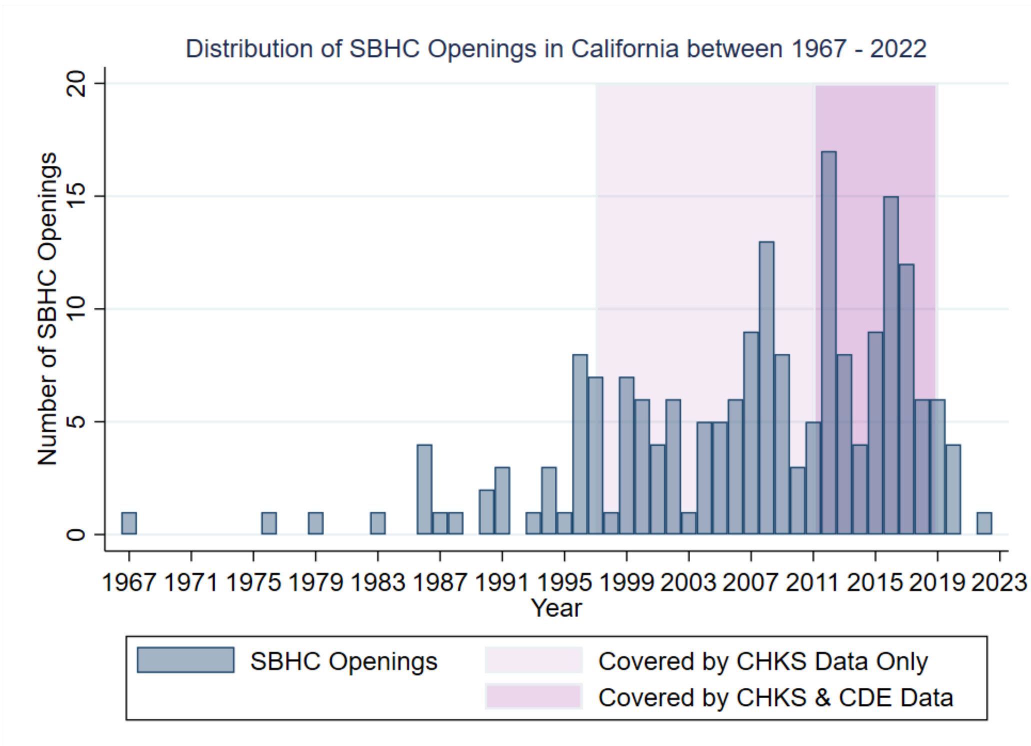

Figure 1: This graph shows the distribution of SBHC openings between 1967-2023 for the set of SBHCs that were still active in California as of 2022. The x-axis shows the range of opening years while the y-axis shows the total number of SBHCs that opened in that year. The light pink shaded area represents the set of SBHC-opening years that can be connected to data from the California Healthy Kids Survey, while the dark purple shaded area marks the set of SBHCs that can be connected to both CHKS data and CDE data on suspensions and dropout rates.

and the name of the SBHC, as well as text fields for the street address, zip code, city, and county.[12] The data do not, however, contain any standardized information (such as a County-District-School (CDS) code, which is used by the CDE to uniquely identify schools in California) or the exact geographic address of the school. In order to connect the CDE outcomes data to the data on SBHC openings, each SBHC opening must first be assigned a valid CDS code.

CDS codes are assigned to schools through an iterative process of fuzzy string matching between the SBHC dataset and the CDE's dataset on"Public Schools and Districts", which contains detailed geographic information and basic school characteristics (such as school name, district and grades offered) for every California public school. The set of potential matches for each SBHC is generated by calculating similarity scores between the *complete address* for each SBHC and each California public school.[13] For every SBHC, each school in the CDE score is assigned a similarity score between 0 and 1, with 1 representing an exact string match. After generating the set of all possible matches for each SBHC, I calculate similarity scores for school name, gradespan, city, county, and zip code within each potential match.[14] Once similarity scores are generated, a "best match" is selected for each SBHC based on nine potential "tiers" of matching criteria.[15] 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. Appendix Table 19 outlines the nine tiers of matching.

Once each SBHC is assigned a principal school, CDE data on suspensions, expulsions, dropout rates, and other demographic school-level data (such as free and reduced price lunch status) are merged on using the county-district-school (CDS) codes. These same CDS codes allow for the merging of the SBHC openings data onto the California Healthy Kids survey data, which is maintained by the CDE. Within the matched sample 73% of SBHCs are on-site, 13% are "off-site" or "telehealth-only", and 14% are mobile vans. Table 1 shows the shares of all SBHCs, just on-site SBHCs, and off-site SBHCs/mobile vans that satisfy a set of characteristics relating to opening dates, gradespans of the primary linked school, categories of populations served, and categories of services offered. This table provides evidence of non-trivial differences between on-site and off-site SBHCs. First, 83% of on-site SBHCs offer mental health services, compared only 48% of off-site SBHCs. On-site SBHCs are also less likely to report serving youth and community members beyond

[12] While a field for the "school name" does exist in the data, this is insufficient for assigning a principal school to each SBHC for two reasons. First, there are a number of cases where schools with the same name exist in different school districts. For example "Jefferson High School" and "Lincoln High School" are common names in this dataset. Since the SBHC data do not have a "school district" field, these cases are difficult to address. Second, since this data is collected directly from SBHCs, there is no standard convention for formatting the name of a school; therefore the same school may be identified by different names in the SBHC data and the official CDE data.

[13] Mechanically, similarity scores are generated with the *matchit* package in Stata which decomposes the text into bigrams before calculating a Jaccard similarity index.

[14] The "gradespan" field is a standardized field in the CDE data, but does not exist in the SBHC openings data; therefore, in order to generate a similarity score on 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".

[15] 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 identfied for each SBHC.

their primary school (although a sizable fraction still do, and therefore there is reason to expect that there might be cross-school spillovers from the opening of an SBHC). A final interesting observation is that over 80% of on-site SBHCs are sponsored either by a Community Health Center (CHC) or a school system. Comparatively, there are a wider range of sponsoring organizations for off-site health centers. This suggests that the factors motivating the opening of an on-site SBHC may be more homogenous than those motivating the opening of an off-site SBHC. These observations lay the foundation for the sample restrictions made in my primary analysis, which are discussed further in the next section.

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

	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

Standard deviations in brackets

3.3 Restricting the Sample of SBHCs

Prior to selecting any control schools, two key restrictions are placed on the sample of school-based health centers considered in this analysis:

1. The SBHC opening year is after 2011 and before 2019
2. The SBHC is defined as "on-site"

The restriction to SBHCs that opened before 2019 addresses the concern that the adverse effects of the COVID-19 pandemic on mental health may attenuate measures of mental-health linked outcomes in the years after 2019. Moreover, this pre-empts the concern that a school that chooses to open an SBHC during or following the pandemic may not necessarily be comparable to a school that opened an SBHC prior to the pandemic. The need to further restrict the sample to SBHCs that opened after 2011 comes from the availability of data on suspensions and dropout rates, both of which are unavailable before the 2011-2012 school year.

Limiting the sample to "on-site" SBHCs leads to the clearest and most relevant interpretation of results for the research question of interest in this paper for two primary reasons. First, compared to off-site and mobile SBHCs, on-site SBHCs are more likely to treat only students in a single school. Figure 2 compares the reported service populations for the three types of SBHCs. On-site SBHCs (illustrated by the red bar) are more likely to report only serving their principal school and less likely to report serving the entire district or multiple districts. Second, limiting to on-site SBHCs provides an obvious "treated group" for the SBHC. If school-based health centers improve take-up of mental health services by reducing *distance* to access services, then those effects should be strongest and most prominent for students in the school that hosts the SBHC. With off-site and mobile vans, the connection to a single "treated school" is less clear and services are likely to be less consistently provided to a single group versus other groups.

One final assumption in the pre-matching sample relates to the definition of an "opening year" for each school-based health center. In order to connect SBHC-opening events with academic outcomes, each SBHC opening must be assigned to an academic year, which covers halves of two calendar years. The data on school-based health centers includes the calendar day, month, and year of the opening, which leaves the researcher to make a choice regarding how these calendar 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 that opened 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

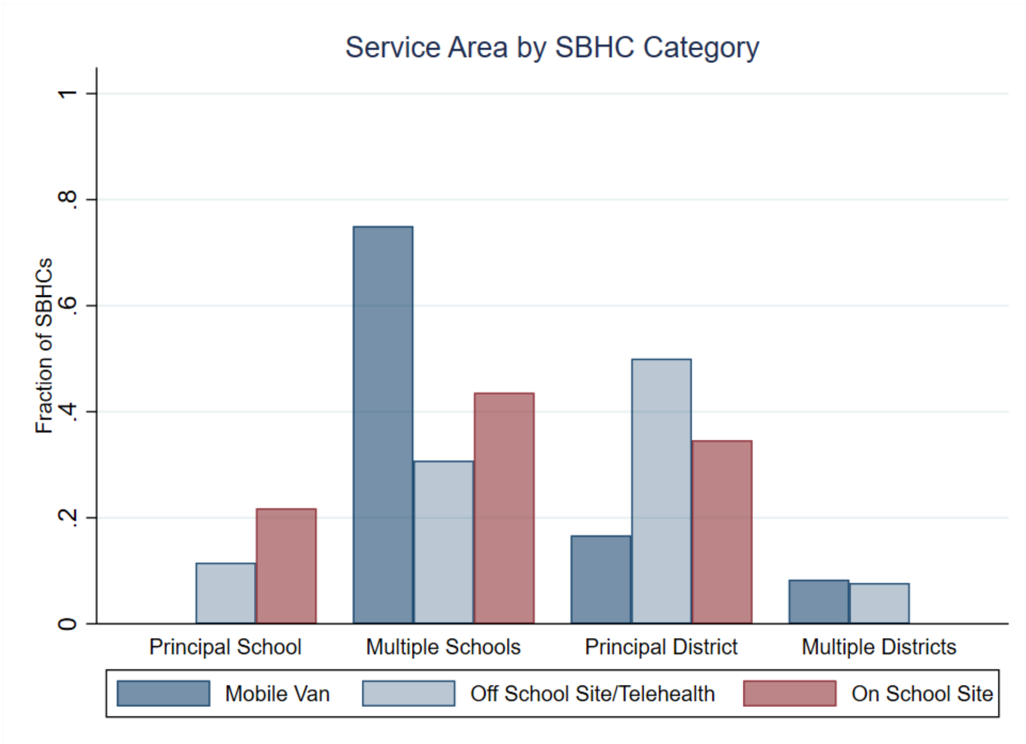

Figure 2: The figure graphs the fraction of school-based health centers within each center "type", that report serving each of the following areas: the principal school (i.e. the school to which the SBHC is attached), multiple schools, the entire district, or multiple districts. Data on the schools served 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 principle district are classified as serving the "principal district"; and SBHCs that list multiple districts are classified as serving "multiple districts".

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 very start of the school year, in the main event study specifications, I weight all post-opening years by the fraction of the year for which the SBHC was open.

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".[16] 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:[17]

1. Caring Staff-Student Relationships
2. School Connectedness
3. Delinquency
4. Substance Use at School

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

[16] https://calschls.org/about/the-surveys/

[17] 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.

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 D.2 describes the data cleaning and index construction process for the CHKS outcomes, as well as the specific questions included in each index.

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:[18]

1. During the past 12 months, did you ever seriously consider attempting suicide? **(Yes/No)**
2. 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.[19] 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.[20] 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

[18] These specific questions draw from other surveys such as the Youth Behavioral Risk Factor Survey administered by the Center for Disease Control and Prevention.

[19] This data can be downloaded from the CDE website's downloadable data files page (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*.

[20] Observations for years after 2019 are dropped due to the confounding COVID-19 pandemic.

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 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.[21] 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.

Table 2 compares the means of each defined outcomes for 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. The primary takeaway from this table is that while dropout rates look similar for both samples, suspension rates are significantly different between schools that ever have an SBHC and schools that never have an SBHC. This 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 and motivates a more thoughtful choice of control group.

4 Empirical Design

The primary empirical design uses a staggered-treatment difference-in-differences model, run both as an event study and a standard two period difference-in-differences regression. This design is most commonly used in contexts where a single policy is "rolled out" to different units (i.e. schools, districts, etc.) at different times. Although the opening of school-based health centers does not originate from a centralized policy rollout, the "staggered" timings of school-based health center openings makes the staggered-treatment difference-in-differences a viable approach. As with any difference-in-differences regression, the validity of the estimates comes from the assumption that outcomes for the treated and control groups would have evolved in parallel in absence of the

[21] Note that for gender or race category **Y** and gradespan **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.

Table 2: Summary Statistics: Suspension Rates and Dropout Rates (Pre-treatment Only)

	Untreated	Treated	p-value
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]	

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.

treatment (this is commonly known as the "Parallel Trends Assumption"). The parallel trends assumption may be violated if control schools experience some deviation in unobserved variables that treated schools do not across the time-period of the study. The standard method of testing this assumption is to check for parallel trends in the years prior to treatment. However, this test should be combined with strong intuition that the parallel trends assumption *should be satisfied*. For example, if the pre-trend does not capture the expected post-event trend for the treated schools in the absence of treatment, the test for parallel trends in the pre-period could be met even with a theoretically "bad" control group.

The primary threat to the parallel trends assumption in this context is that the decision to open an SBHC and the timing of that decision are not unconditionally random with respect to district and school-level trends; however, the *exact timing* of the opening may be plausibly exogenous. The non-randomness of the decision to open an SBHC comes from the observation that the decision to open a school-based health center often comes from a confluence of sources, including school administrators, school district leaders, and sometimes community health organizations. The potential exogeneity of the exact timing of an opening comes from the observation that the standard timeline for constructing an SBHC can take around 2-3 years and 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.

The presence of selection into opening a school-based health center must be accounted for in the empirical approach. In policy-rollout contexts where selection is not a concern, the traditional choices for control groups are either: *future-treated units* (under the assumption that the time of treatment is random and that units treated at different times will be similar on unobservable characteristics); *never-treated units* (under the assumption that units that receive treatment are randomly selected and therefore similar to never-treated units on unobservables); or a combination of both. In the presence of selection, the set of all future-treated and never-treated units may not be appropriate controls. Appendix B shows evidence that the results from my primary specifications do not hold under either of these alternate choices of control group. In order to address selection, I employ a propensity-score matching model to construct a matched control group for the set of school-based health center openings. With the use of a propensity-score matched control group, the validity of difference-in-differences estimates now relies on conditional parallel trends. Specifically, conditional on having similar predicted likelihoods of opening an SBHC, the outcomes for treated schools would evolve similarly to outcomes in the selected control schools in the absence of treatment. This is likely to be satisfied if, 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.

Propensity score matching has been proposed as a viable quasi-experimental approach to addressing selection bias (Dehejia and Wahba (1999), Dehejia and Wahba (1999)), but has come under scrutiny in recent years. For example Smith and Todd (2005) notes that the results from propensity-score matching models can be highly sensitive to the choice of predictors used to construct the propensity score, the specification of the model, and the choice of alternate control groups. Results from Heckman et al. (1997) emphasize the importance of propensity score matching within the same "local labor market" (i.e. a school-district or local education agency in the education context) and measuring the dependent variable in the same way across treated and control schools in order to mitigate un-observable factors that may bias the similarity of matches; however Smith and Todd (2005) notes that while these geographic restrictions are important, they can be reasonably relaxed when the propensity score matching is combined with a difference-in-differences estimation. This important insight from Smith and Todd (2005) 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.

Matching locally (i.e. within the same school-district) poses two concerns in this setting. The first is that finding a good within-district match is not always possible outside of large school districts (such as Los Angeles Unified, for example). 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* houses 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. Appendix B shows that within-district propensity score matching in fact does not lead to a control group with similar pre-trends to the treated schools. Given this results, I relax the requirement of matching within-district; however in an effort to reduce bias from un-observable factors I place certain restrictions on the set of districts from which controls are selected. For a school with an SBHC opening in year y I restrict the set of districts from which a match is selected to districts that *have at least one SBHC* that opened within 5 years of the first opening in the sample and up to year y. This omits districts that opened an SBHC much earlier than the study window and districts that opened an SBHC after the specific treated school. [22] In this way, the sample of potential districts have a similar "openness" or underlying propensity to support an SBHC within a few years of the treated school.[23]

In the final matching process, a school that opens an SBHC in year y is matched to a control school that: (1) has the same gradespan (i.e. elementary, middle, or high school); (2) comes from a district that is "open" to having SBHC; and (3) has the closest predicted propensity of having an SBHC in year y. The "propensity score" measures the likelihood of opening an SBHC in school s in year y, and uses as predictors the lagged fraction of Free and Reduced Price Meal (FRPM) students (a standard proxy for socioeconomic status) and the lagged enrollment at a school (a proxy for school

[22]Note that the reason for omitting districts that opened an SBHC after the specific treated school follows the same logic as avoiding using future-treated schools as controls for a currently treated school.

[23]Note that this approach allows for a match to be made within the same district if this is the *best possible match* but does not force in-district matches and therefore reduces that concern of "bad matches".

size). The process of selecting these predictors and the functional form of the propensity score is discussed 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.

This matching procedure generates a sample of 44 SBHCs with matched control schools (yielding a total analysis sample of 88 schools). Of these 44 schools, only 3 have a "best match" that is within the same district. Table 3 shows the balance of a vector of school-level characteristics measured in the pre-event period, 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 significant 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. A first takeaway from Column (1) is that the two variables used in matching (fraction of FRPM students and total enrollment) are balanced across the treated and control schools. Looking to variables that may be correlated with the decision to open an SBHC but are not use in matching, there is a slight imbalance in the fraction of URM students between treated and control schools; however the magnitude of the difference amounts to only 6% of the control group mean. Regardless, to control for any residual imbalances, my preferred specification controls for these school-level characteristics. Table 4 shows summary statistics for all outcomes in the pre-event period, for the entire matched sample.

Table 3: Difference in Sample Means ($Control - Treatment$)

	All Matched	Elementary	Middle	High
	Diff [Contr μ]	Diff [Contr μ]	Diff [Contr μ]	Diff [Contr μ]
Fraction FRPM	0.031 [0.719]	0.006 [0.908]	0.021 [0.647]	0.066 [0.695]
	(0.226)	(0.746)	(0.580)	(0.145)
Total Enrollment	24.333 [1210.534]	81.299 [639.325]	-31.234 [1577.500]	141.833 [950.714]
	(0.783)	(0.062)	(0.817)	(0.070)
Fraction URM	0.06** [0.80]	-0.03 [0.90]	0.10** [0.78]	0.04 [0.76]
	(0.010)	(0.214)	(0.003)	(0.365)
N	348	78	186	84

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. * p<0.05, ** p<0.01, *** p<0.001

Table 4: Summary Statistics: Suspension Rates and Dropout Rates (Pre-treatment Only)

	Mean	St. Dev.	Min	Max	N
Suspension Outcomes					
Suspension Rate	0.057	0.057	0.000	0.303	227
Female Suspension Rate	0.037	0.042	0.000	0.249	227
Male Suspension Rate	0.077	0.074	0.000	0.451	227
Defiance-Only Suspension Rate	0.021	0.037	0.000	0.248	227
Non-defiance Suspension Rate	0.036	0.031	0.000	0.169	227
Violence Suspension Rate	0.047	0.054	0.000	0.312	227
Weapon Possession Suspension Rate	0.002	0.003	0.000	0.019	227
Illicit Drug Suspension Rate	0.012	0.015	0.000	0.070	227
Dropout Outcomes					
Dropout Rate	0.012	0.015	0.000	0.075	149
Female Dropout Rate	0.010	0.013	0.000	0.080	149
Male Dropout Rate	0.014	0.017	0.000	0.074	149

Once treatment-control pairs are identified, the following event-study specification is estimated:

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

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^τ is a dummy equal to 1 if the observation is τ years after (or before if τ is negative) the opening year for its matched pair. ω_τ 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 ω_τ 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". γ_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 that accounts for the presence of an unbalanced panel:

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

where Y_{st} and $Treated_s$ are defined as above. γ_s is a set of school fixed effects and ν_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). $\mathbb{X}_{st}$ is the same vector of school characteristics defined above. As before, standard errors are clustered at the school level.

5 Results

5.1 Mental Health Correlations

One novel contribution of this study comes from the use of data from the California Healthy Kids Survey (CHKS). Through a partnership with the California Department of Education I acquired access annual survey data from 1998-2021. An ideal analysis using this data would connect the universe of SBHC openings in California with survey data from the CHKS in order to directly assess the impact of a school-based health center opening on measures of school climate and mental health. This analysis requires: (a) a sufficient number of schools with an SBHC and matched control schools represented in the data; (b) a sufficient number of pre-opening and post-opening years of survey data to validate the parallel trends assumption and test for treatment effects. The manner and frequency of CHKS survey administration makes the satisfaction of both of these assumptions difficult. First, districts select into participation in the survey and districts that choose to administer the survey usually do so for a sub-set of schools in the district. As a result, not every school in California is represented in the CHKS data. 64% of schools with an SBHC and 23% of schools without an SBHC have at least one year of CHKS data. Secondly, most districts that administer the CHKS do so on a biannual basis. As a result, less than 22% of schools with an SBHC in the data have both: (1) a minimum of two years of CHKS data before and after the SBHC opening; and (2) a maximum gap of two years between consecutive data collections.

Given these data limitations, a difference-in-differences or event-study model is likely to be highly underpowered; however this data may still be useful in evaluating the mental health impacts of SBHC openings. Table 5 presents simple regressions of suspension rates on each of the following four CHKS indices measuring school climate: level of reported delinquency, level of reported substance use, the presence of caring staff-student relationships, and level of school connectedness. 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. Columns (1) - (4) of Table 5 show that suspension rates are positively correlated with student-reported levels of delinquency and substance use. In a similar vein, 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.

Table 5: Correlations Between School Climate and Suspension Rates

	Suspension Rates							
	(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
Delinquency (1-5)	0.094***	0.021***						
	(0.007)	(0.004)						
Substance Use (1-5)			0.037***	0.007***				
			(0.005)	(0.003)				
Caring Staff-Student (1-5)					-0.062***	-0.009***		
					(0.003)	(0.003)		
School Connectedness (1-5)							-0.062***	-0.015***
							(0.002)	(0.002)
Constant	0.017***	0.070***	0.042***	0.078***	0.268***	0.112***	0.305***	0.138***
	(0.006)	(0.004)	(0.006)	(0.004)	(0.008)	(0.009)	(0.008)	(0.008)
Observations	10878	10436	10881	10438	10882	10441	10882	10441

Standard errors in parentheses

Observations are at the school level.

* $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 6 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.

Table 6: Correlation Between Mental Health and Suspension Rates

	Suspension Rates			
	(1)	(2)	(3)	(4)
	Year FE	Year/School FE	Year FE	Year/School FE
Fraction of Students Considered Suicide	0.032**	0.011		
	(0.013)	(0.020)		
Fraction of Students Experienced Depression			0.044***	0.014**
			(0.006)	(0.007)
Constant	0.078***	0.081***	0.054***	0.062***
	(0.004)	(0.005)	(0.002)	(0.002)
Observations	4809	3759	8819	8348

Standard errors in parentheses

Observations are at the school level.

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

Tables 7 and 8 show regressions of dropout rates on the same set of school climate measures and mental health measures respectively. Examining Table 7, a first observation is that the magnitude of the relationships between measures of school climate and dropout rates is much smaller than 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.

Looking instead at mental health measures in Table 8, 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.

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 negative correlation between suspension rates and school climate measures that reflect how comfortable and interconnected 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.

Table 7: Correlations Between School Climate and Dropout Rates

	Dropout Rates							
	(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
Delinquency (1-5)	0.020***	0.004**						
	(0.002)	(0.002)						
Substance Use (1-5)			0.008***	0.000				
			(0.001)	(0.001)				
Caring Staff-Student Relationships (1-5)					-0.008***	0.004***		
					(0.002)	(0.001)		
School Connectedness (1-5)							-0.011***	0.003**
							(0.001)	(0.001)
Constant	-0.003	0.010***	0.003	0.013***	0.036***	0.003	0.050***	0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.005)	(0.004)	(0.003)	(0.004)
Observations	8169	7761	8172	7765	8173	7767	8175	7769

Standard errors in parentheses

Observations are at the school level.

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

Table 8: Correlations Between Mental Health and Dropout Rates

	Dropout Rates			
	(1)	(2)	(3)	(4)
	Year FE	Year/School FE	Year FE	Year/School FE
Fraction of Students Considered Suicide	-0.003	0.004		
	(0.006)	(0.004)		
Fraction of Students Experienced Depression			0.021***	-0.008
			(0.005)	(0.007)
Constant	0.020***	0.018***	0.000	0.009***
	(0.002)	(0.001)	(0.001)	(0.002)
Observations	2865	2642	5271	4752

Standard errors in parentheses

Observations are at the school level.

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

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

The first three columns of Table 9 show the estimated coefficients from a set of event study specifications following Equation 1, where the dependent variable is the school-level suspension rate. Specification (1) is a simple two-way fixed effects (TWFE) specification with school fixed-effects, event-time dummies, a treatment dummy, and interaction between treatment and each of the event-time dummies. Specification (2) adds a vector of school-level controls, that includes the fraction of FRPM students, fraction of minority students, and total enrollment. This is my 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). Specification (3) adds linear time trends for each gradespan[24] to test for the robustness of these results to gradespan-specific changes over time in the outcome. Columns (4) - (6) show three analogous specifications applying the alternate model from Callaway and Sant'Anna (2021) that accounts for the possibility of bias in staggered-event difference-in-differences arising from heterogenous treatment effects across treated cohorts or time periods.[25] Section 6.1 provides further discussion of whether this bias is likely to occur in this sample.

All columns of Table 9 show an insignificant and close to 0 effect of the treatment in the pre-event years, indicating 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 point 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 a p-value of around 0.35. 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. Section 6.2 reveals that increasing the sample size by expanding from 1:1 propensity-score matching to 3:1 matching (up to three control schools for each treated

[24]Mechanically, Specification (3) adds separate fixed effects for elementary schools, middle schools, and high schools, each interacted with a continuous variable for calendar year

[25]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.

school) yields treatment effects of a similar magnitude for years 1-3 that are now individually and jointly significant at the 5% level (with an F-test p-value of 0.035 for my preferred specification). Comparing the Callaway & Sant'Anna specification from Column (5) to the corresponding two-way fixed effect model in Column (2), the adjustment for heterogeneous treatment effects leads to coefficients for event-years 1-2 that are smaller in magnitude than the TWFE coefficients, but a similar magnitude coefficient 3 years following the opening. A χ^2-test that the coefficients on years 0-3 are jointly equal to zero yields a p-value of 0.307, suggesting that the significance of these coefficients is not meaningfully different from that of the original TWFE coefficients.

Figures 3 and 4 present two methods of visualizing the results in Column (2). In Figure 3 I plot separate event studies for the treated and control groups.[26] This allows for a visual verification that: (1) the parallel trends assumption holds; and (2) that any differences between the treated and control groups over time is driven by treatment group deviations from the pre-event trend rather than control group deviations. Figure 3 shows a nearly perfect match on pre-trends for the three years prior to treatment. There is also a visible deviation in trend after year 0 for the treated group, suggesting that any estimated treatment effects are in fact driven by changes in treated schools.

Figure 4 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 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.12 shows the corresponding event-study graph for the Callaway & Sant'Anna adjusted regression from Column (5) of Table 9.

In order to estimate the average effect an SBHC opening in the post-opening period, I run difference-in-differences versions of the three event study regressions from Table 9. Columns (1)-(3) of Table 10 show two-way fixed effects regressions while columns (4)-(6) show the Callaway & Sant'Anna equivalents.[27] The preferred TWFE specification in column (2) shows that for treated schools, suspension rates decrease by an average of 1.1 percentage points in the years after an SBHC opening. This 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 4.2%, this amounts to a 27% decrease, from the expected baseline rate. The preferred Callaway & Sant'Anna specification in Column (5) estimates a treatment effect of the same magnitude that is similarly significant at the 5% level.

[26]The specific regression here is a simple modification of Equation 1 that omits the $Treated_s$ dummy and the interaction between $Treated_s$ and the D_t dummies.

[27]For columns (4) - (6) 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 estimate should be interpreted as the average treatment effect across all schools that opened an SBHC in any year between 2012-2019.

Table 9: Event Study: Suspensions Rates

	Standard TWFE			Callaway & Sant'Anna		
	(1)	(2)	(3)	(4)	(5)	(6)
Treated x (τ = -3)	0.0024	0.0026	0.0022	-0.0004	0.0001	-0.0015
	(0.0094)	(0.0094)	(0.0093)	(0.0082)	(0.0087)	(0.0092)
Treated x (τ = -2)	-0.0007	-0.0005	-0.0007	-0.0046	-0.0054	-0.0035
	(0.0074)	(0.0075)	(0.0075)	(0.0057)	(0.0068)	(0.0078)
Treated x (τ = -1)	ref.	ref.	ref.	ref.	ref.	ref.
Treated x (τ = 0)	-0.0011	-0.0013	-0.0012	-0.0076*	-0.0054	-0.0065
	(0.0069)	(0.0071)	(0.0071)	(0.0046)	(0.0055)	(0.0068)
Treated x (τ = 1)	-0.012	-0.012	-0.012	-0.011	-0.008	-0.010
	(0.009)	(0.008)	(0.008)	(0.007)	(0.007)	(0.009)
Treated x (τ = 2)	-0.014	-0.015	-0.015	-0.012	-0.009	-0.009
	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.012)
Treated x (τ = 3)	-0.013	-0.012	-0.012	-0.015	-0.013	-0.022*
	(0.008)	(0.008)	(0.008)	(0.010)	(0.010)	(0.012)
Baseline Treatment Effect	0.049***	0.057	-0.098			
	(0.003)	(0.072)	(2.789)			
School FE	Yes	Yes	Yes	Yes	Yes	Yes
School Characteristics	No	Yes	Yes	No	Yes	No
Gradespan Time Trends	No	No	Yes	No	No	Yes
F-Stat/Chi-Stat	1.067	1.107	1.110	8.990	4.815	5.989
p-value	0.373	0.353	0.351	0.061	0.307	0.200
Pre-Period Control Mean	0.042	0.042	0.042	0.042	0.042	0.042
R^2	0.837	0.837	0.839			
Observations	505	505	505	505	505	453

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.

Treated x (τ = -1) represents the omitted reference category.

* $p<0.1$, ** $p<0.05$, *** $p<0.01$

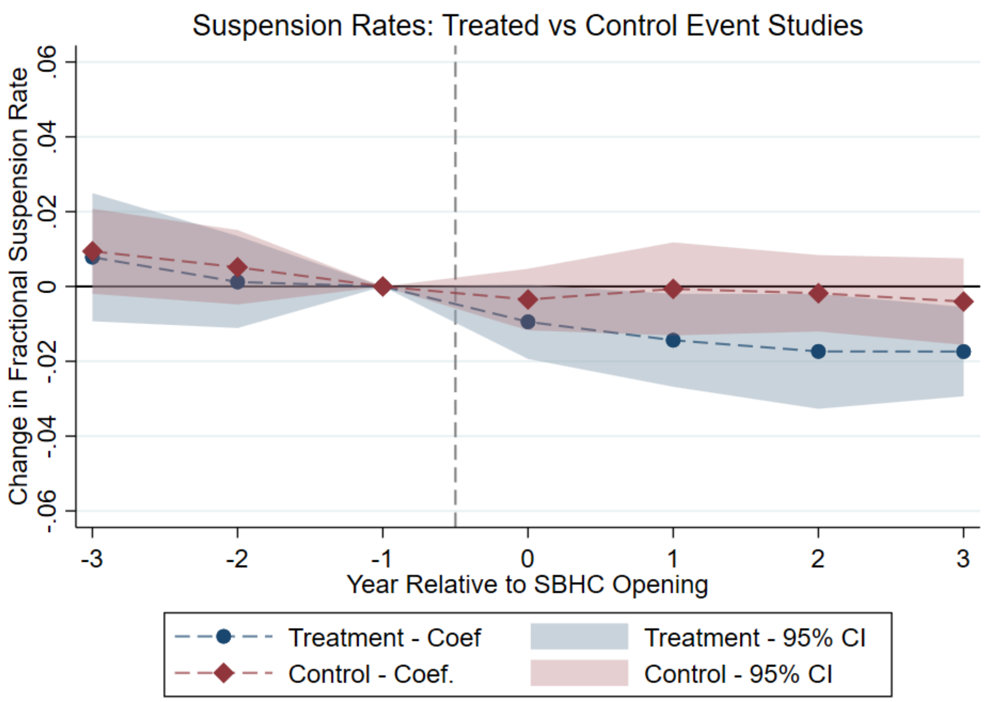

Figure 3: This figure plots the *Event Time* coefficients from separate event studies restricting the sample to all treated schools and all control schools respectively for the outcome of suspension rates. Both sub-sample event studies control for school fixed effects and a vector of school characteristics that includes fraction of Free and Reduced Price Meal (FRPM) students, fraction of under-represented 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.

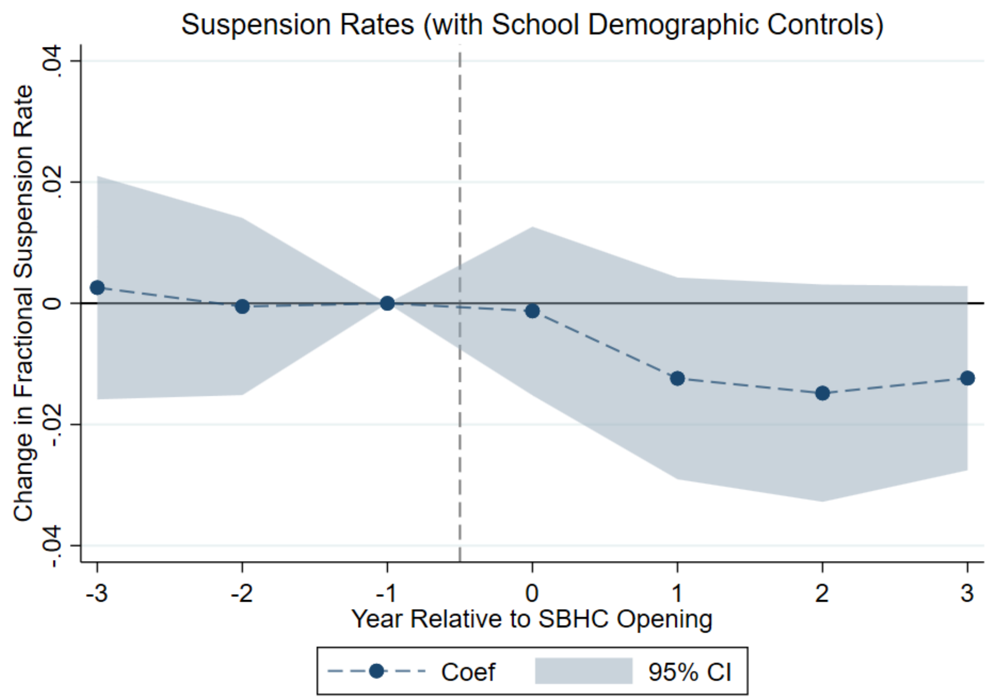

Figure 4: This figure plots the $Treatment \times Event\ Time$ coefficients from an augmented event study, controlling for school fixed effects and a vector of school characteristics that includes fraction of Free and Reduced Price Meal (FRPM) students, fraction of underrepresented minority students, 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.

Table 10: Difference-in-Differences: Suspension Rates

	Standard TWFE			Callaway & Sant'Anna		
	(1)	(2)	(3)	(4)	(5)	(6)
Treated X Post	-0.0113**	-0.0114**	-0.0110**	-0.0123**	-0.0107**	-0.0130*
	(0.0056)	(0.0056)	(0.0054)	(0.0051)	(0.0051)	(0.0068)
Constant	0.0625***	0.0961	-0.5280			
	(0.0046)	(0.0593)	(0.4947)			
School FE	Yes	Yes	Yes	-	-	-
Year FE	Yes	Yes	Yes	-	-	-
School Characteristics	No	Yes	Yes	No	Yes	Yes
Gradespan Time Trends	No	No	Yes	No	No	Yes
Pre-Period Control Mean	0.0419	0.0419	0.0419	0.0419	0.0419	0.0419
R^2	0.841	0.841	0.844			
Observations	505	505	505	505	505	453

Standard errors in parentheses. Observations are at the school level. * p<0.1, ** p<0.05, *** p<0.01
The *csdid* package does not estimate the constant. The inclusion of school and year fixed effects is not relevant for the Callaway and Sant'Anna method, as it does not use a two-way fixed effects approach.

5.3 Dropout Rates

Following the same structure as the previous section, Columns (1) - (3) of Table 11 show the results from three event-study specifications where the dependent variable is now the school-level dropout rate. Columns (4) and (5) show the models from the baseline specification (Column 1) and the preferred specification (Column 2), but using the alternate estimation approach from Callaway and Sant'Anna (2021). Due to insufficient sample size, I am unable to properly estimate the specification from Column (3) using the Callaway and Sant'Anna estimator.[28] As before, Columns (2) and (5) show the results from my preferred specification. 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 11 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 1.5 percentage points in year 3 that is statistically significant at the 5% level. Compared to the control school baseline dropout rate of 1.2% this is a nearly 125% decrease. Applying the Callaway & Sant'Anna adjustment in Column (5) leads to a smaller estimate that is no longer statistically significant, suggesting that this coefficient should be interpreted with caution.

[28] This is not a significant concern, as Column (3) is only included as a robustness check.

Table 11: Event Study: Dropout Rates

	Standard TWFE			Callaway & Sant'Anna	
	(1)	(2)	(3)	(4)	(5)
Treated x (τ = -3)	-0.0015	-0.0010	-0.0007	0.0002	0.0007
	(0.0037)	(0.0033)	(0.0033)	(0.0026)	(0.0029)
Treated x (τ = -2)	0.0019	0.0015	0.0017	0.0022	0.0018
	(0.0038)	(0.0037)	(0.0037)	(0.0029)	(0.0033)
Treated x (τ = -1)	ref.	ref.	ref.	ref.	ref.
Treated x (τ = 0)	0.0004	0.0008	0.0010	0.0013	0.0002
	(0.0035)	(0.0036)	(0.0036)	(0.0020)	(0.0022)
Treated x (τ = 1)	-0.000	0.000	-0.001	0.001	0.000
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Treated x (τ = 2)	-0.002	-0.001	-0.002	0.003	0.002
	(0.005)	(0.006)	(0.006)	(0.003)	(0.005)
Treated x (τ = 3)	-0.016***	-0.015***	-0.015***	-0.011***	-0.011
	(0.003)	(0.003)	(0.004)	(0.004)	(0.008)
Baseline Treatment Effect	-0.000	-0.026	1.964		
	(0.002)	(0.022)	(1.782)		
School FE	Yes	Yes	Yes	Yes	Yes
School Characteristics	No	Yes	Yes	No	Yes
Gradespan Time Trends	No	No	Yes	No	No
F-Stat/Chi-Stat	8.879	13.224	12.238	8.841	2.166
p-value	0.000	0.000	0.000	0.065	0.705
Pre-Period Control Mean	0.012	0.012	0.012	0.012	0.012
R^2	0.860	0.871	0.874		
Observations	240	240	240	234	234

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.

Treated x (τ = -1) represents the omitted reference category.

In Column (6) a insufficient sample size prevents estimation of the coefficient for Treated x (τ = 3).

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

Figures 5 and 6 once again show two ways of visualizing these results. Figure 5 plots separate event studies for the treated and control groups, while Figure 6 shows the standard event study plot. In both figures, there is no evidence of a violation of the parallel pre-trends assumption. Both figures also show a clear decrease in dropout rates for treated schools three years following the opening, but no effect in years 0-2. As noted before, this effect becomes insignificant under the Callaway and Sant'Anna (2021) adjusted model (see Appendix Figure A.13).

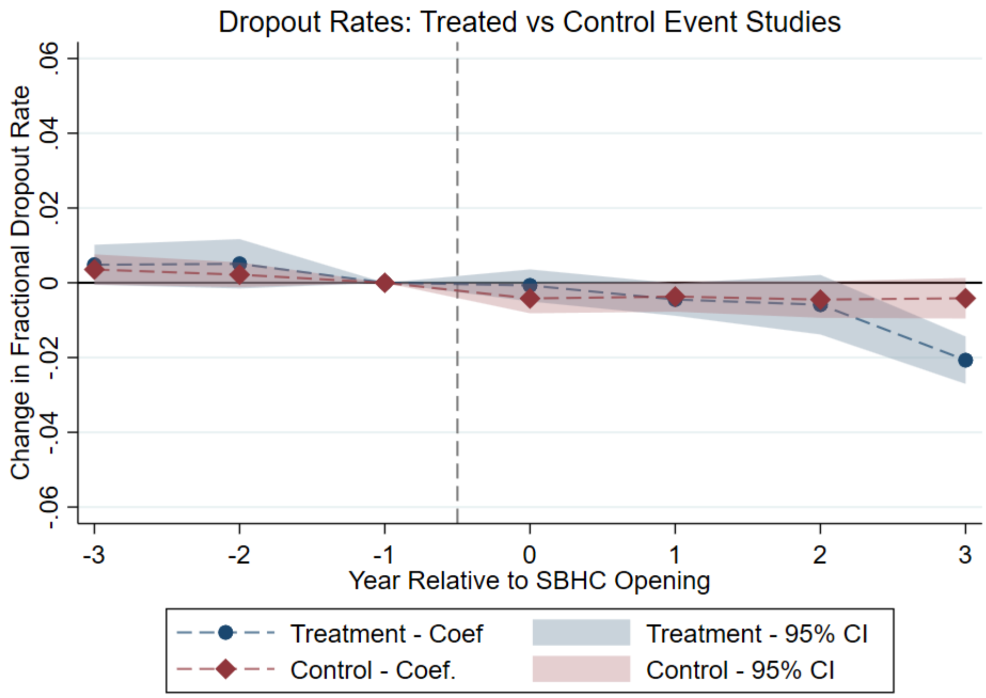

Figure 5: This figure plots the *Event Time* coefficients from separate event studies restricting the sample to all treated schools and all control schools respectively, for the outcome of dropout rates. Both sub-sample event studies control for school fixed effects and a vector of school characteristics that includes fraction of Free and Reduced Price Meal (FRPM) students, fraction of under-represented 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.

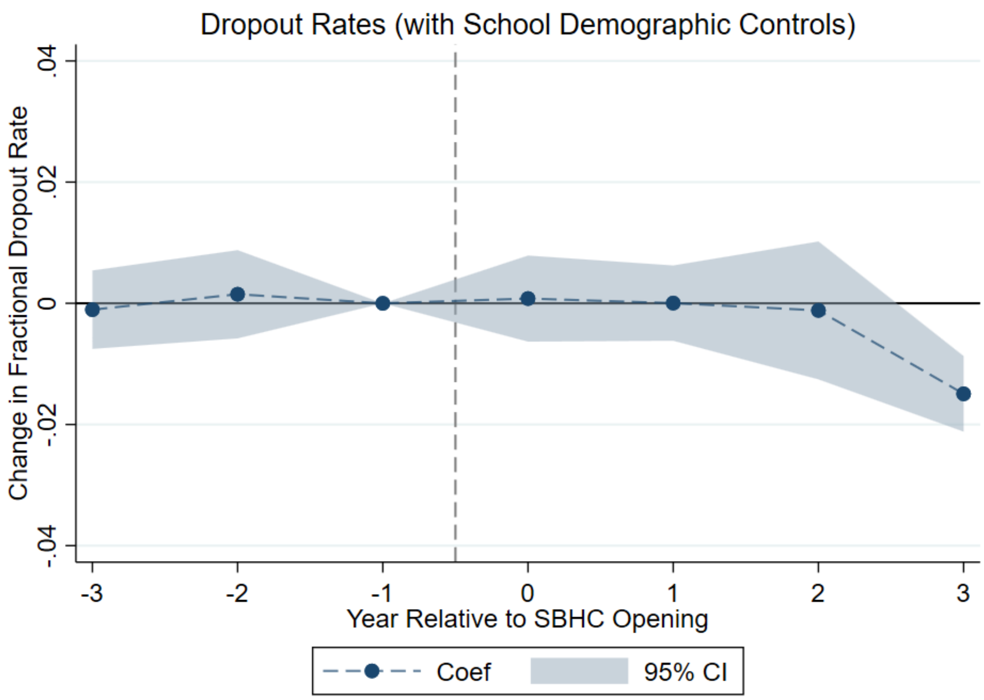

Figure 6: This figure plots the $Treatment \times Event\ Time$ coefficients from an augmented event study for dropout rates, controlling for matched pair fixed effects and a vector of school characteristics that includes fraction of Free and Reduced Price Meal (FRPM) students, fraction of underrepresented 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.

Finally, Table 12 shows the corresponding difference-in-differences estimates.[29] My preferred specifications for both the two-way fixed-effects (Column 2) and Callaway & Sant'Anna (Column 5) models show no average change in dropout rates for schools with an SBHC after the SBHC-opening. The difference-in-differences estimate fails to capture the drop three years after the event that is visible in the event studies; however it suggests that a conservative estimate of the average treatment effect on dropout rates would be zero. 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 the dropout rate of greater than 0.5 of a percentage point. Given that the estimated decrease 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. As before columns (4)-(6) show that these effects are not sensitive to bias from heterogeeous treatment effects across groups or time periods.

Table 12: Difference-in-Differences: Dropout Rates

	Standard TWFE			Callaway & Sant'Anna	
	(1)	(2)	(3)	(4)	(5)
Treated X Post	-0.0003	0.0000	-0.0006	-0.0000	-0.0007
	(0.0023)	(0.0023)	(0.0023)	(0.0013)	(0.0015)
Constant	0.0101***	-0.0239	1.8331		
	(0.0032)	(0.0198)	(1.5190)		
School FE	Yes	Yes	Yes	-	-
Year FE	Yes	Yes	Yes	-	-
School Characteristics	No	Yes	Yes	No	Yes
Gradespan Time Trends	No	No	Yes	No	No
Pre-Period Control Mean	0.0119	0.0119	0.0119	0.0119	0.0119
R^2	0.859	0.867	0.870		
Observations	240	240	240	234	234

Standard errors in parentheses. Observations are at the school level. * p<0.1, ** p<0.05, *** p<0.01
The *csdid* package does not estimate the constant. The inclusion of school and year fixed effects is not relevant for the Callaway and Sant'Anna method, as it does not use a two-way fixed effects approach.

[29] Note that once again, the Callaway and Sant'Anna adjustment for Column (3) of Table 12 is omitted due to sample size limitations.

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. In Section 5 I supplement my standard TWFE regressions with results from the alternate estimator proposed by Callaway and Sant'Anna (2021) to show that my main results are largely robust to treatment effect heterogeneity. 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.[30]

[30]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.

In this section I attempt to further bolster the claim that bias from heterogeneous treatment 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.[31] 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.

6.2 k-Nearest Neighbors Matching Adjustment

The main analysis sample in this paper uses a one-to-one propensity-score matching process where each SBHC is matched to a single control school with the closest propensity score. One question that may arise here is whether these results are exclusive to the small sample of treatments and controls selected. In order to test this, I run a k-neighbor matching variation of the original propensity score matching process. Specifically, I hold constant all aspects of the matching process (including the maximum acceptable distance between propensity scores for a "good" match, the propensity score model, and all restrictions placed on the pool of potential control schools) and allow for up to 3 matches to be selected for each treated school. This substantially increases the size of the analysis sample; therefore if the results are not driven by a select set of schools, we would expect the results to be of a similar magnitude to the original estimates, but with a more precisely estimated for the control schools. There is also, however, the possibility that expanding the matching process to allow for up to three matches will worsen the match on pre-trends. In particular, if the first control selected is the one with the "most similar" propensity score, each subsequent control has an increased chance of being a further distance from the best match. Therefore, expanding to a k-nearest neighbors matching process has the potential to lead to a set of control schools that is a worse match for the treated schools on average.

[31] In practice, this is implemented using the *twowayfeweights* package in Stata.

Table 13 shows event study estimates for the three specifications used throughout this paper, on the sample of matched pairs with up to three matches per treated school. Looking first at the pre-event coefficients, while the coefficient magnitudes are slightly larger than in the one-to-one matching specification, there is still no identifiable treatment effect prior to the SBHC opening. The coefficients on years 1-3 following the opening reveal larger magnitude decreases (of around 1.5 - 1.7 percentage points) that are statistically significant at the 5% level.

Figure 7 shows separate event studies for treated and control schools in this sample. A key takeaway here is that although the increase in sample size leads to a more precisely estimated trend for control units, the pre-trends become more dissimilar with this sample. This is a primary reason that the main sample selected for this paper prioritizes one-to-one nearest neighbor matching rather than k-nearest neighbor matching.

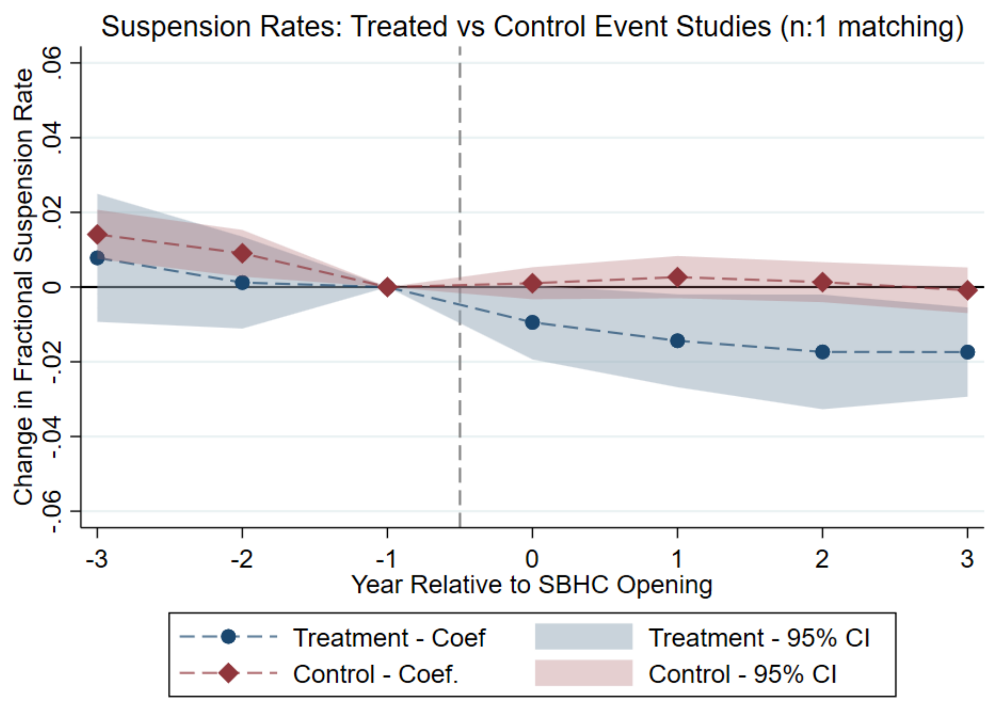

Figure 7: This figure plots the *Event Time* coefficients from separate event studies restricting the sample to all treated schools and all control schools respectively for the outcome of suspension rates. The sample used comes from a 3-nearest neighbors matching model. Both sub-sample event studies 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.

Table 14 shows the corresponding event study specifications for the outcome of dropout rates. Once again, this reveals a similar post-event trend to the original specification, with no effect 1-2 years following treatment, and a decrease of around 1.1 percentage points three years following treat-

Table 13: Event Study: Suspensions Rates (n:1 matching)

	(1)	(2)	(3)
Treated x (τ = - 3)	-0.0043	-0.0036	-0.0036
	(0.0083)	(0.0083)	(0.0083)
Treated x (τ = - 2)	-0.0063	-0.0054	-0.0055
	(0.0065)	(0.0064)	(0.0064)
Treated x (τ = -1)	ref.	ref.	ref.
Treated x (τ = 0)	-0.009	-0.008	-0.008
	(0.006)	(0.006)	(0.006)
Treated x (τ = 1)	-0.016**	-0.015**	-0.015**
	(0.007)	(0.007)	(0.007)
Treated x (τ = 2)	-0.017**	-0.017**	-0.016**
	(0.008)	(0.008)	(0.008)
Treated x (τ = 3)	-0.016**	-0.015**	-0.015**
	(0.007)	(0.007)	(0.006)
Baseline Treatment Effect	0.160***	0.204	-0.309
	(0.007)	(0.136)	(0.763)
Fraction FRPM		0.006	0.001
		(0.013)	(0.014)
Fraction Minority		-0.041	-0.067
		(0.051)	(0.061)
School Size		0.000**	0.000
		(0.000)	(0.000)
School FE	Yes	Yes	Yes
School-Level Time Trends	No	No	Yes
F-Stat	1.640	1.503	1.598
p-value	0.162	0.199	0.173
Pre-Period Control Mean	0.041	0.041	0.041
R^2	0.834	0.835	0.837
Observations	991	991	991

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.

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

ment. Figure 8 shows the separate event studies. Once again, there seems to be more dissimilarity in the pre-trend compared to the one-to-one matching model, although this is less distinguishable for dropout rates due to the low magnitude of the estimates.

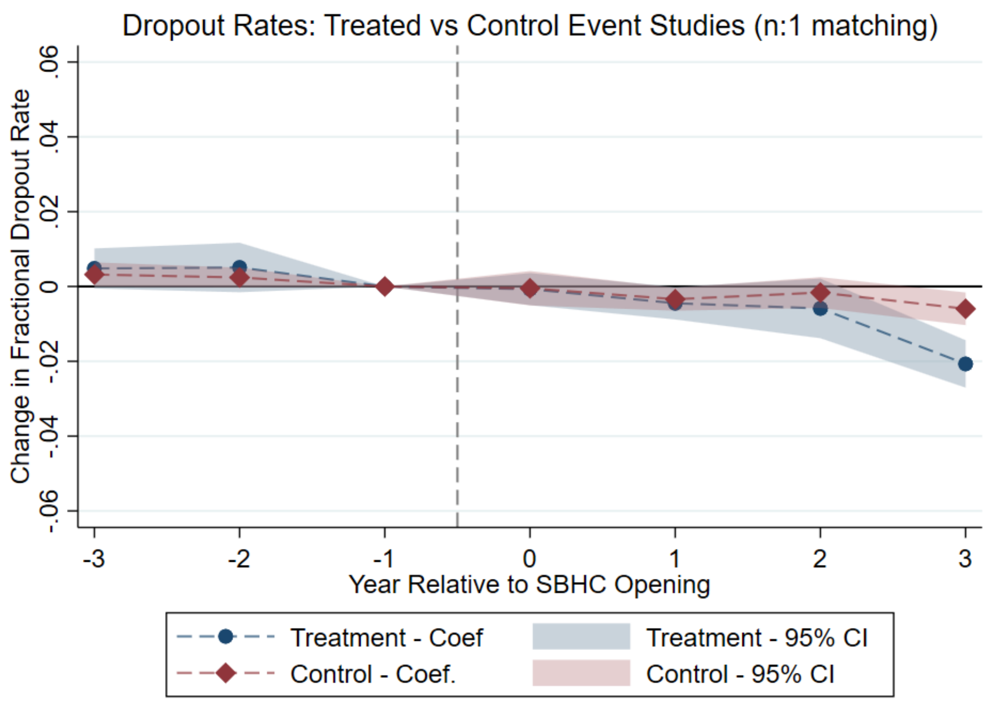

Figure 8: This figure plots the *Event Time* coefficients from separate event studies restricting the sample to all treated schools and all control schools respectively for the outcome of dropout rates. The sample used comes from a 3-nearest neighbors matching model. Both sub-sample event studies control 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.

Table 14: **Table 14:** Event Study: Dropout Rates (n:1 matching)

	(1)	(2)	(3)
Treated x (τ = - 3)	0.0007	0.0009	0.0010
	(0.0031)	(0.0031)	(0.0031)
Treated x (τ = - 2)	0.0019	0.0022	0.0022
	(0.0033)	(0.0033)	(0.0033)
Treated x (τ = -1)	ref.	ref.	ref.
Treated x (τ = 0)	-0.002	-0.002	-0.002
	(0.004)	(0.004)	(0.004)
Treated x (τ = 1)	-0.001	-0.000	-0.000
	(0.003)	(0.003)	(0.003)
Treated x (τ = 2)	-0.003	-0.002	-0.002
	(0.005)	(0.005)	(0.005)
Treated x (τ = 3)	-0.013***	-0.011***	-0.011***
	(0.003)	(0.003)	(0.004)
Baseline Treatment Effect	0.043***	0.075	-0.082
	(0.003)	(0.063)	(0.966)
Fraction FRPM		-0.005	-0.003
		(0.016)	(0.016)
Fraction Minority		-0.016	-0.005
		(0.034)	(0.039)
School Size		0.000	0.000
		(0.000)	(0.000)
School FE	Yes	Yes	Yes
School-Level Time Trends	No	No	Yes
F-Stat	4.562	4.313	3.835
p-value	0.001	0.002	0.005
Pre-Period Control Mean	0.012	0.012	0.012
R^2	0.815	0.816	0.817
Observations	492	492	492

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.

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

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 first by looking at heterogeneity of treatment effects by gender and then by suspension type. 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 9 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.

Table 15 shows the point-estimates from the original pooled sample event study in Column (1), followed by the sub-sample estimates for male and female students in Columns (2) and (3) respectively. These estimates show a decrease in suspension rates of between 1.5 - 2.4 percentage points (with the coefficient of -0.024 in event year 2 reaching marginal significance at the 10% level). Comparatively, for females, the suspension rate decreases by around 1 percentage point one year following the opening, but only by 0.6 of a percentage point in the years that follow.

Table 16 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 1.5 percentage points, while the corresponding decrease for female students in only 0.7 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.

Beyond just identifying which groups are most affected by the opening of an SBHC, we may also be interested in understanding whether the decreases in suspensions is driven by one or more specific categories of offense. In particular, 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 treatable mental health issues. The California Department of

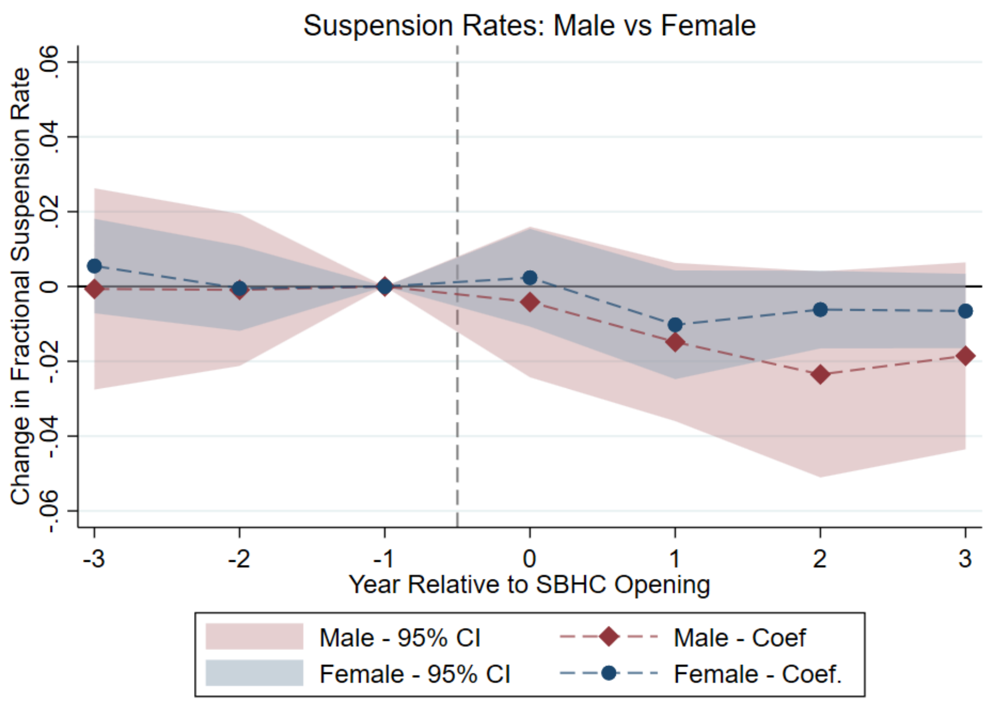

Figure 9: This figure plots the *Event Time* coefficients from separate event studies for the outcomes of male suspension rate **(red)** and female suspension rates **(blue)**. Both sub-sample event studies control for school fixed effects and a vector of school characteristics that includes fraction 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.

Table 15: Suspensions Rates: Gender Heterogeneity

	(1) Pooled	(2) Male	(3) Female
Treated x (τ = - 3)	0.0026	-0.0007	0.0057
	(0.0094)	(0.0137)	(0.0065)
Treated x (τ = - 2)	-0.0005	-0.0009	-0.0003
	(0.0075)	(0.0104)	(0.0059)
Treated x (τ = -1)	ref.	ref.	ref.
Treated x (τ = 0)	-0.001	-0.004	0.002
	(0.007)	(0.010)	(0.007)
Treated x (τ = 1)	-0.012	-0.015	-0.010
	(0.008)	(0.011)	(0.007)
Treated x (τ = 2)	-0.015	-0.024*	-0.006
	(0.009)	(0.014)	(0.005)
Treated x (τ = 3)	-0.012	-0.019	-0.006
	(0.008)	(0.013)	(0.005)
Baseline Treatment Effect	0.057	0.095	0.015
	(0.072)	(0.092)	(0.054)
Fraction FRPM	-0.000	0.018	-0.019
	(0.026)	(0.033)	(0.023)
Fraction Minority	-0.049	-0.105	0.026
	(0.066)	(0.090)	(0.054)
School Size	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
School FE	Yes	Yes	Yes
F-Stat	1.107	1.186	1.256
p-value	0.353	0.316	0.287
Pre-Period Control Mean	0.042	0.058	0.025
R^2	0.837	0.806	0.832
Observations	505	505	505

τ represents the time with respect to the SBHC-opening year ($\tau = 0$)
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 16: Suspensions Rates: Gender Heterogeneity (DiD)

	(1) Pooled	(2) Male	(3) Female
Treated X Post	-0.0114**	-0.0153*	-0.0074**
	(0.0056)	(0.0083)	(0.0037)
Fraction FRPM	0.0004	0.0165	-0.0180
	(0.0244)	(0.0321)	(0.0218)
Fraction Minority	-0.0365	-0.0912	0.0293
	(0.0528)	(0.0731)	(0.0418)
School Size	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
Constant	0.096	0.151**	0.029
	(0.059)	(0.074)	(0.049)
School FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Pre-Period Control Mean	0.0419	0.0584	0.0246
R^2	0.841	0.811	0.833
Observations	505	505	505

Standard errors in parentheses

Observations are at the school level.

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

Education defines six categories of offenses that can lead to a suspension: violent incident (with injury), violent incident (no injury), weapons possession, illicit drug-related, defiance-only, and other.[32] To isolate suspensions resulting from treatable 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".

Figure 10 shows the event study regression results for defiance-only suspensions and non-defiance suspensions (which includes suspensions in the five other offense categories). Note that the outcome for defiance suspensions is the fraction of *total students enrolled* who were suspended for defiance reasons, while the outcome for non-defiance suspension is the fraction of total students enrolled who were suspended for any reason that does not qualify as "defiance only". The primary takeaway from this figure is that there is a clear decline in defiance-only suspensions beginning in the year of the SBHC opening. Comparatively there seems to be no effect of the opening on non-defiance suspensions.

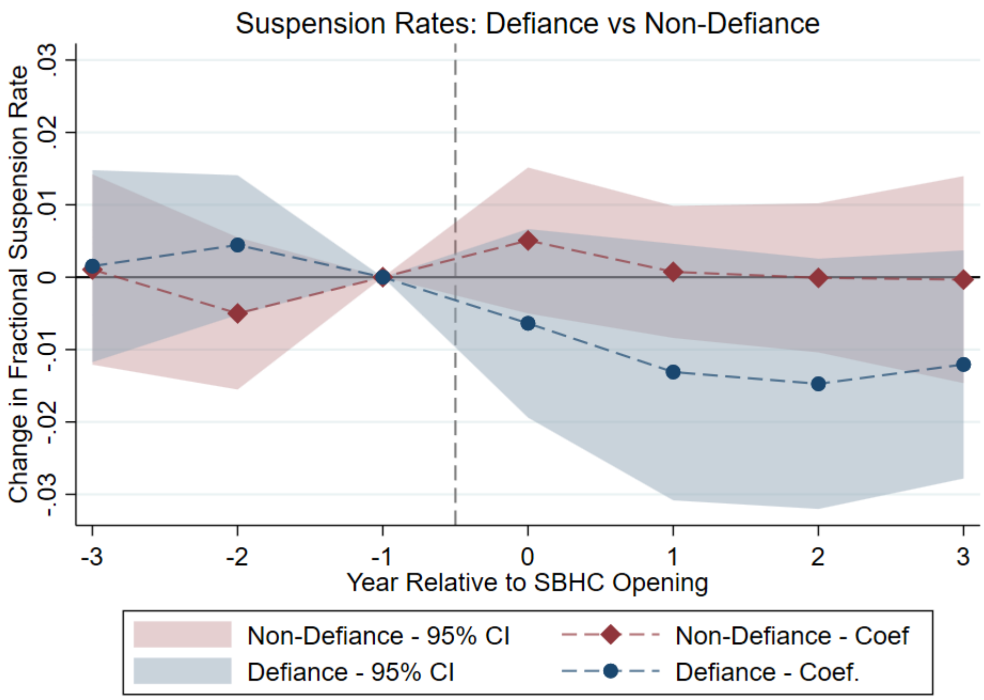

Figure 10: This figure plots the *Event Time* coefficients from separate event studies where the outcomes are: the non-defiance suspension rate **(red)** and the defiance-only suspension rate suspension rates **(blue)**. Both sub-sample event studies control for school fixed effects and a vector of school 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.

Table 17 shows the estimated event study coefficients for overall suspension rates (Column 1), the

[32]The specific offenses included in each of these categories are outlined in Appendix Section B.

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. For defiance-only suspensions, the coefficients for 1-3 years following the SBHC opening almost exactly parallel the coefficients for the overall suspension rate. Comparatively, for non-defiance suspensions, the treatment effects are effectively 0 for all post-opening years. 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). These results indicate that the entirety of the decline in suspensions may be driven by a decrease in disruptive behavior, which suggests that the decrease in suspensions may be due to direct improvements in treatment of behavioral and mental health disorders rather than a crowd-out of suspensions through other avenues. For example, if there were a significant decrease in suspensions caused by weapons possession as a result of the SBHC opening, this may raise concerns that the opening of an SBHC is concurrent with changes in other school-level policies (such as security measures or severity of policing) that might directly impact the rate of weapons possession.

Table 18 shows difference-in-differences estimates for the overall, defiance-only, and non-defiance suspension rates in Columns (1) - (3). Once again, Column (2) reveals a statistically significant decrease in suspension rates of a similar magnitude to the estimate for the overall suspension rate (around 1.18 percentage points), with no significant treatment effect for all non-defiance suspensions (Column 3). To further investigate whether the effect for non-defiance suspensions is truly zero, Columns (4) - (6) show the difference-in-differences estimates for three sub-categories of non-defiance offenses: violence suspensions, weapons possession suspensions, and illicit drug possession suspensions. The coefficients on the interaction between treatment and the years after an SBHC-opening suggest that if there is a decrease in non-defiance suspensions, it is caused by decreases in illicit drug use suspensions (which has an estimated decrease of 0.17 of a percentage points); however this estimate is statistically insignificant and the magnitude is less than one-fifth of the overall decrease in suspension rates, suggesting once again, that this category is not a primary driver of the treatment effect on suspension rates. It is not, however, surprising that there could be an effect on illicit drug use. In particular, if students' use of illicit substances is in response to depression, poor mental health, or other health issues, an SBHC has the potential to address those root issues. That the largest decrease is for suspensions caused by disruptive behavior suggests that the types of mental illnesses or behavioral issues that cause disruptive actions may be the issues that are most treatable by the limited capacity and services of a school-based health center.

Table 17: Suspensions Rates: Defiance vs Non-Defiance Heterogeneity

	(1) All Offenses	(2) Defiance-Only	(3) Non-Defiance
Treated x (τ = - 3)	0.0026	0.0015	0.0011
	(0.0094)	(0.0068)	(0.0067)
Treated x (τ = - 2)	-0.0005	0.0044	-0.0050
	(0.0075)	(0.0049)	(0.0054)
Treated x (τ = -1)	ref.	ref.	ref.
Treated x (τ = 0)	-0.001	-0.006	0.005
	(0.007)	(0.007)	(0.005)
Treated x (τ = 1)	-0.012	-0.013	0.001
	(0.008)	(0.009)	(0.005)
Treated x (τ = 2)	-0.015	-0.015*	-0.000
	(0.009)	(0.009)	(0.005)
Treated x (τ = 3)	-0.012	-0.012	-0.000
	(0.008)	(0.008)	(0.007)
Baseline Treatment Effect	0.057	0.084	-0.026
	(0.072)	(0.067)	(0.040)
Fraction FRPM	-0.000	0.003	-0.004
	(0.026)	(0.018)	(0.018)
Fraction URM	-0.049	-0.099	0.050
	(0.066)	(0.093)	(0.042)
School Size	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)
School FE	Yes	Yes	Yes
F-Stat	1.107	0.769	0.325
p-value	0.353	0.546	0.861
Pre-Period Control Mean	0.042	0.014	0.028
R^2	0.837	0.648	0.805
Observations	505	505	505

τ represents the time with respect to the SBHC-opening year ($\tau = 0$)

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 18: Suspensions Rates: Heterogeneity by Offense Type (DiD)

	(1) Any Offense	(2) Defiance-Only	(3) Non-Defiance (All)	(4) Violence	(5) Weapon Poss.	(6) Illicit Drug
Treated X Post	-0.0114**	-0.0118*	0.0003	-0.0084	0.0002	-0.0017
	(0.0056)	(0.0062)	(0.0037)	(0.0052)	(0.0004)	(0.0018)
Fraction FRPM	0.0004	0.0102	-0.0102	-0.0071	0.0013	-0.0117
	(0.0244)	(0.0188)	(0.0190)	(0.0267)	(0.0018)	(0.0110)
Fraction URM	-0.0365	-0.0875	0.0512	0.0262	-0.0007	0.0248*
	(0.0528)	(0.0740)	(0.0424)	(0.0490)	(0.0034)	(0.0149)
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	0.096	0.099	-0.003	0.031	0.002	-0.000
	(0.059)	(0.072)	(0.040)	(0.044)	(0.004)	(0.016)
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Pre-Period Control Mean	0.0419	0.0141	0.0277	0.0329	0.0015	0.0082
R^2	0.841	0.649	0.808	0.840	0.537	0.710
Observations	505	505	505	505	505	505

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 that is concentrated primarily amongst "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 may 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 0.9 - 1.1 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 4.2%, so this predicted range of effects represents a 21% - 27% decrease in suspension rates. Subgroup analyses reveal that while both male and female students are significantly affected, the decrease in suspensions is larger for males. Finally, I provide suggestive evidence that this 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 that the drop is driven almost entirely by a decrease in suspensions for disruptive behavior, which is often the result of untreated mental health issues. This combination of results points to a positive impact of school-based health centers on addressing mental health issues that may manifest in behavioral issues.

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

(2022). Oakland, CA.

(2023). Income eligibility scales for school year 2023–24 - rates, eligibility scales, & funding (ca dept of education).

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.

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).

CSBHA (2023). Funding school-based health & wellness centers: California school-based health alliance.

CSBHA, C. S.-B. H. A. (2022). Navigating the promise of sbhcs: A guide for health care leaders.

CSHCA (2010).

de Chaisemartin, C. and D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96.

Dehejia, R. H. and Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American statistical Association*, 94(448):1053–1062.

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.

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.

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-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*.

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 school-based 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.

A Tables & Figures

Table 19: Matching criteria for nine iterations of fuzzy 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")

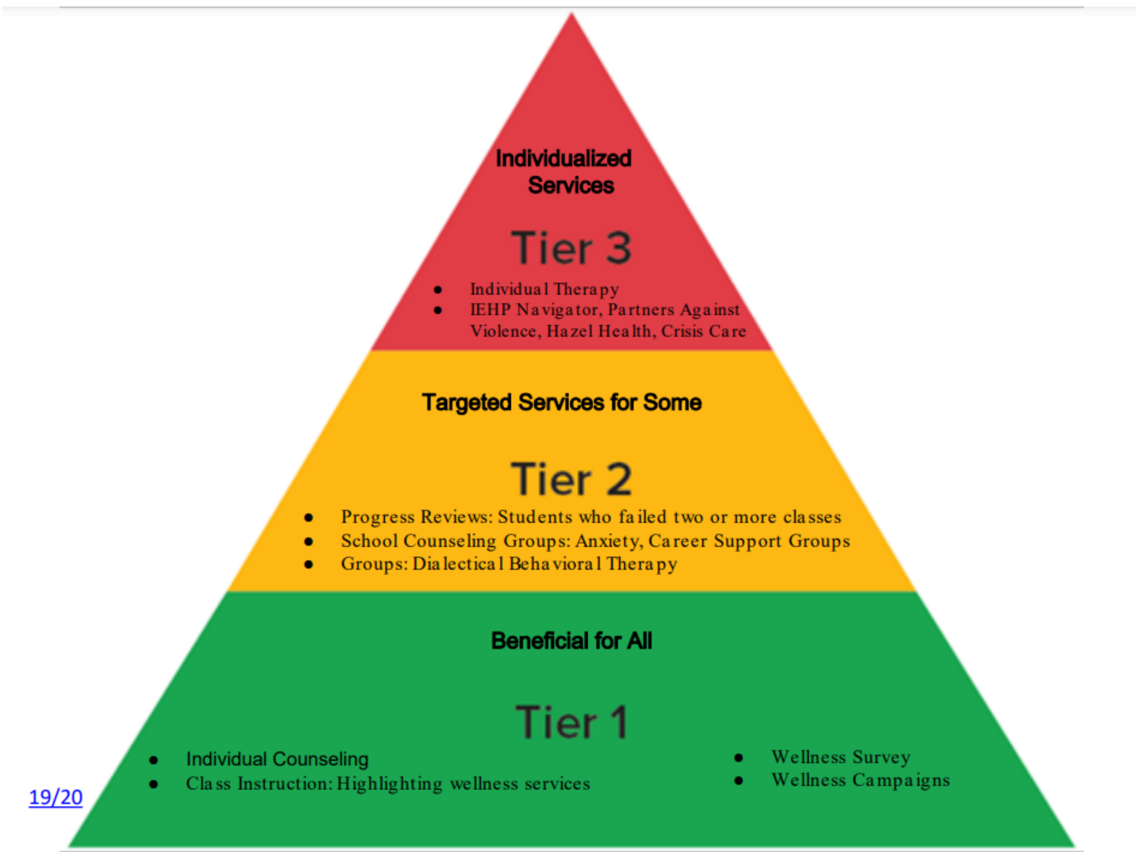

Figure A.11: This figure comes from a presentation delivered by the California School-Based Health Alliance (CBSHA) at their annual conference in 2023. It outlines the tiers of mental-health service provision recommended for new SBHCs by the CBSHA. Tier 1 services are services that all SBHCs in California 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.

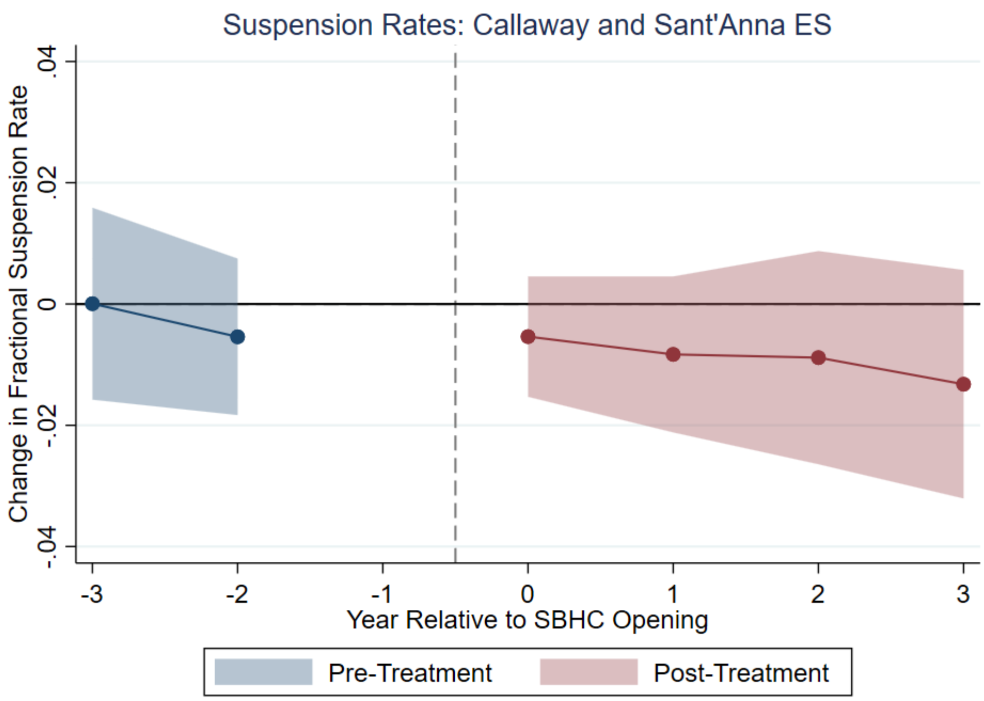

Figure A.12: This figure shows the event study plot from a Callaway and Sant'Anna (2021) regression where the outcome is the suspension rate. The specification includes a vector of school characteristic controls whose baseline values are used for the inverse propensity-score weighting used to aggregate individual 2×2 difference-in-differences ATT estimates up to the event-time level.

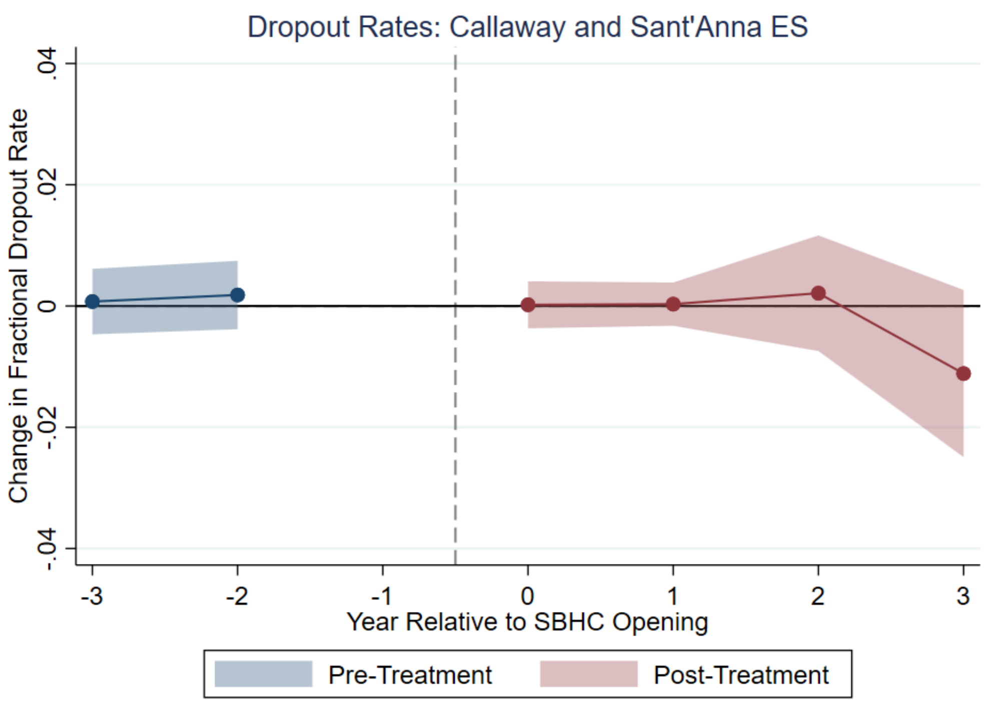

Figure A.13: This figure shows the event study plot from a Callaway and Sant'Anna (2021) regression where the outcome is the dropout rate. The specification includes a vector of school characteristic controls whose baseline values are used for the inverse propensity-score weighting used to aggregate individual 2×2 difference-in-differences ATT estimates up to the event-time level.

B Propensity Score 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, 2022). While socioeconomic status could be measured through median income, this data is not available at the school level.[33] 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.[34] 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.[35] 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 "on-site" 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.

[33] The most granular geography for Census data on median household income is the census tract level.

[34] 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. (a20, 2023)

[35] 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. $\mathbb{X}_{s,t-1}$ is a vector of lagged predictors for school s in year $t-1$. In the most saturated specification $\mathbb{X}_{s,t-1}$ contains 1-year lags of the fraction of students qualified for FRPM, the fraction minority students, and the total enrollment. Table 20 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 χ^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 20 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).

Table 20: Predicted Likelihood of Having an SBHC - Variable Selection

	(1)	(2)	(3)
1Y FRPM Lag	1.215***	1.760***	2.075***
	(0.246)	(0.257)	(0.343)
1Y Enrollment Lag		0.00105***	0.00106***
		(0.0000627)	(0.0000635)
1Y Minority Lag			-0.438
			(0.354)
Constant	-4.895***	-6.160***	-6.047***
	(0.199)	(0.212)	(0.234)
Chi Squared	24.32	342.9	341.7
Observations	18243	18189	18188

Standard errors in parentheses

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

Once the correct predictors have been determined, it remains to select the correct function of these predictors. Table 21 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 χ^2 statistic for the specification in Column (2) is lower than the χ^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.

Table 21: Predicted Likelihood of Having an SBHC - Model Selection

	(1)	(2)	(3)	(4)
1Y FRPM Lag	1.760***	1.725***	2.125	2.043
	(0.257)	(0.254)	(1.448)	(1.439)
1Y Enrollment Lag	0.00105***	0.00167***	0.00105***	0.00166***
	(0.0000627)	(0.000243)	(0.0000643)	(0.000242)
$(1YEnrollmentLag)^2$		-0.000000194***		-0.000000194***
		(7.04e-08)		(7.02e-08)
$(1YFRPMLag)^2$			-0.294	-0.257
			(1.169)	(1.165)
Constant	-6.160***	-6.430***	-6.251***	-6.509***
	(0.212)	(0.235)	(0.397)	(0.411)
Chi Squared	342.9	318.0	351.4	329.6
Observations	18189	18189	18189	18189

Standard errors in parentheses

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

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 21, as the two specifications where all predictors are statistically significant. Table 22 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 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

χ^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 $\mathbb{X}_{s,t-1}$ 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.

Table 22: Predicted Likelihood of Having an SBHC (Logit)

	Elementary		Middle		High	
	(1)	(2)	(3)	(4)	(5)	(6)
1Y FRPM Lag	5.607***	5.780***	0.597*	0.539	2.849***	2.849***
	(0.824)	(0.871)	(0.331)	(0.332)	(0.945)	(0.944)
1Y Enrollment Lag	-0.000102	0.00958***	0.000691***	0.00196***	-0.00106***	-0.000997
	(0.000373)	(0.00281)	(0.0000644)	(0.000303)	(0.000363)	(0.000977)
$(1YEnrollmentLag)^2$		-0.00000866***		-0.000000416***		-4.28e-08
		(0.00000259)		(0.000000101)		(0.000000442)
Constant	-9.370***	-11.92***	-4.282***	-4.812***	-5.130***	-5.151***
	(0.678)	(1.169)	(0.269)	(0.324)	(0.827)	(0.856)
Chi Squared	55.30	57.19	115.4	96.89	16.31	28.04
Observations	10943	10943	4014	4014	2321	2321

Standard errors in parentheses

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 that are treated earlier. This approach relies on the assumption that conditional on two 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.14 shows the separate event studies for treated and control schools using a "future-treated" 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.[36] Figure C.14 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.

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 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

[36] 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.

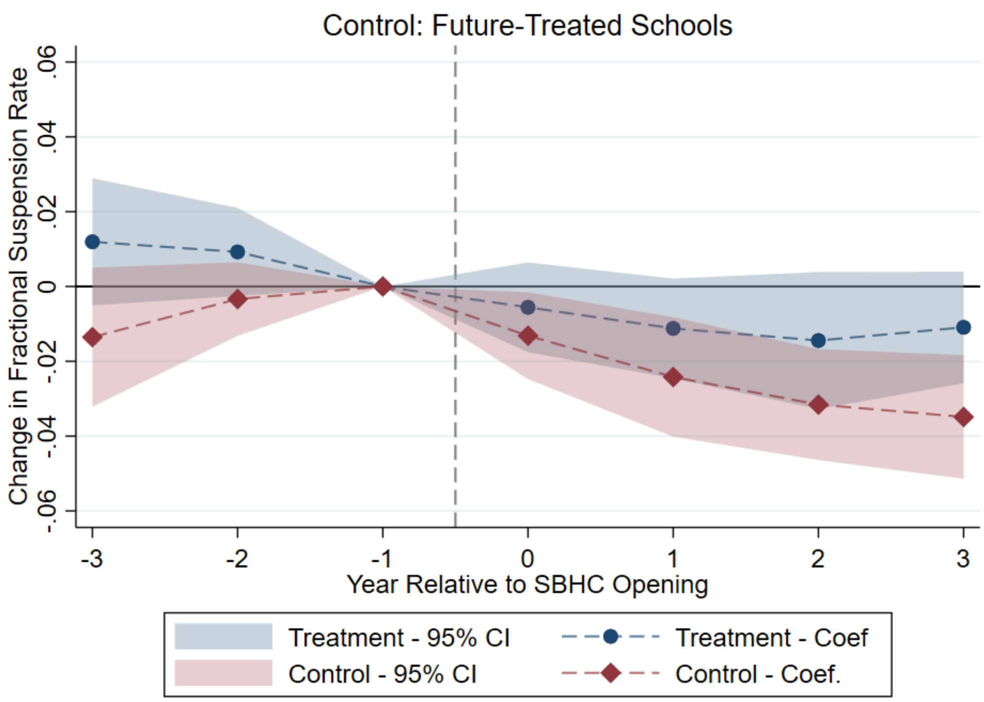

Figure C.14: This figure plots the *Event Time* coefficients from separate treated and control group event studies using *future-treated* schools as controls. Both sub-sample event studies control for school fixed effects and a vector of school characteristics that includes fraction of Free and Reduced 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.

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 $k <= y$. 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.15 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.

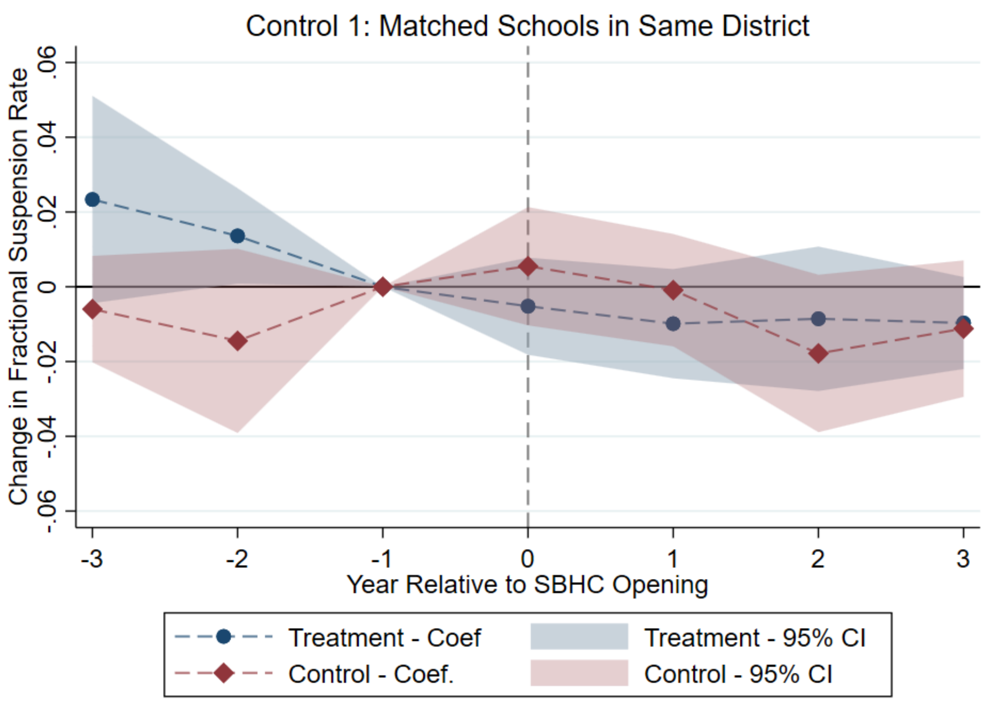

Figure C.15: This figure plots the *Event Time* coefficients from separate treated and control group event studies using control schools that are selected through 1:1 propensity-score matching from *the same school district* as each treated school. Both sub-sample event studies control for school fixed 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.

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 23 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.16 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 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.

Table 23: Summary Statistics: District & School Characteristics (2012 Data)

	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]	

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.

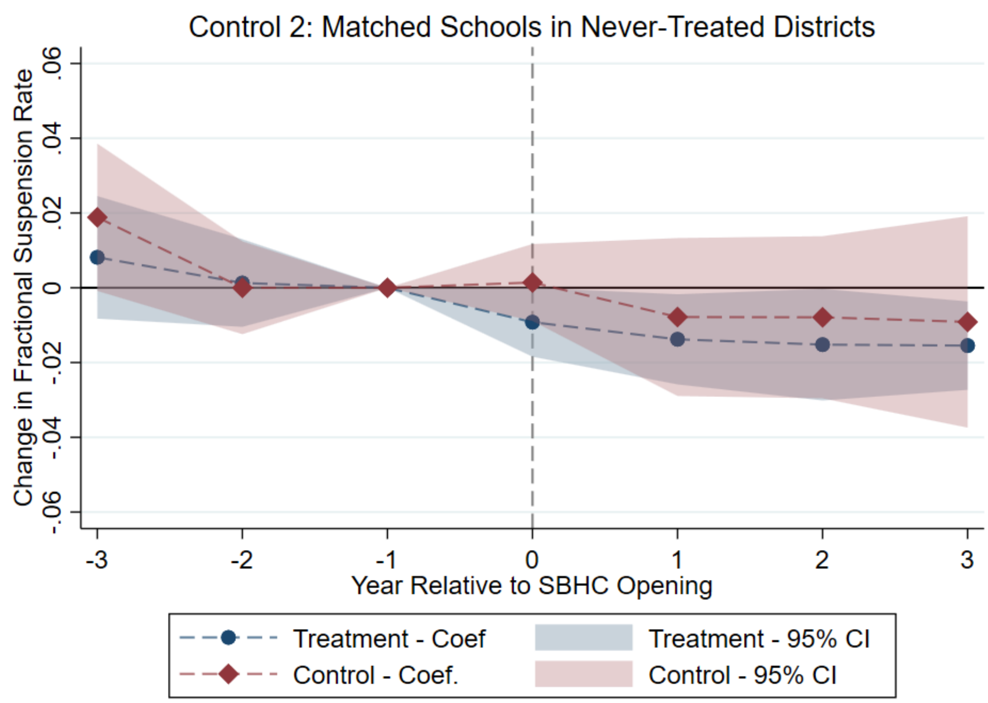

Figure C.16: This figure plots the *Event Time* coefficients from separate treated and control group event studies using control schools that are selected through 1:1 propensity-score matching from *never-treated districts* as each treated school. Both sub-sample event studies control for school fixed 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 California Department of Education Data Descriptions

D.1 Suspension Offense Categories

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 Substance, Alcohol, Intoxicant	48900(c)
	Offering, Arranging, or Negotiating Sale of Controlled Substances, Alcohol, Intoxicants	48900(d)
	Offering, Arranging, or Negotiating Sale of Drug Paraphernalia	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)

Table 24: 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

D.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 suppors 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.[37]. 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 25 partially reproduces a table from Mahecha and Hanson (2020) that lists the questions included in each index and the weight assigned to each question.[38]

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 D.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.

[37] *Source:* https://calschls.org/about/the-surveys/

[38] 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	0.851
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)	0.681
been offered, sold, or given an illegal drug at school (12 months)	0.707
damaged school property on purpose at school (12 months)	0.745
carried a gun at school (12 months)	0.846
carried any other weapon at school (12 months)	0.778
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 days)	0.930
electronic cigarettes, e-cigarette on school property (30 days)	0.864
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 25: 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).