The Call is Coming from Inside the School! How Well Does Cell Phone Data Predict Whether K12 School Buildings Were Open During the Pandemic?

Dan Goldhaber
Nick Huntington-Klein
Nate Brown
Scott Imberman
Katharine O. Strunk

November 2024

WORKING PAPER No. 309-1124

CALDER
National Center for Analysis of Longitudinal Data in Education Research

AIR®

The Call is Coming from Inside the School! How Well Does Cell Phone Data Predict Whether K12 School Buildings Were Open During the Pandemic?

Dan Goldhaber
American Institutes for Research / CALDER
University of Washington

Nick Huntington-Klein
Seattle University

Nate Brown
University of Washington

Scott Imberman
Michigan State University

Katharine O. Strunk
University of Pennsylvania
CALDER

Contents

Acknowledgments

The authors are grateful to SafeGraph, the Michigan Department of Education, the Michigan Center for Educational Performance Information, and the Washington Office of Superintendent of Public Instruction for providing some of the data necessary to conduct this research. We appreciate comments and suggestions from Michael DeArmond, Zeyu Xu. We also thank Rafia Nisat, Daniel Moore, Federico Chung, Suvekshya Gautam, and Bryant Hopkins for excellent research assistance.

Disclaimer: Nick Huntington-Klein formerly provided consulting services to SafeGraph and was paid for his services.

CALDER • American Institutes for Research
1400 Crystal Drive 10th Floor, Arlington, VA 22202
202-403-5796 • www.caldercenter.org

The Call is Coming from Inside the School! How Well Does Cell Phone Data Predict Whether K12 School Buildings Were Open During the Pandemic?

Dan Goldhaber, Nick Huntington-Klein, Nate Brown, Scott Imberman, & Katharine O. Strunk

CALDER Working Paper No. 309-1124
November 2024

Abstract

The COVID-19 pandemic forced widespread school closures and a shift to remote learning. A growing body of research has examined the effects of remote learning on student outcomes. But the accuracy of the school modality measures used in these studies is questionable. The most common measures—based on self-reports or district website information—are often inconsistent and lack nationwide coverage. Some studies have used cell phone mobility data to identify school modalities, but there is no consensus yet on how to translate device pings into modality measures. This paper contributes to the literature on modality measurement by examining the relationship between mobile device signals and school modality prior to the pandemic and applies those findings to the pandemic period in Michigan and Washington. We compare our results to state-provided closure data and other nationwide sources, including the Return to Learn Tracker and the COVID-19 School Data Hub. Our findings indicate that cell phone mobility data can accurately predict school modality under normal conditions, but the accuracy drops during the pandemic. These results have implications for future research on educational and health outcomes during both pandemic and non-pandemic-related school closures.

1. Introduction

The onset of the COVID-19 pandemic in the spring of 2020 led to widespread school building closures across the globe. In many places, schools continued to operate completely or partially remote throughout the 2020-21 school year. While the full implications of remote learning for student and community outcomes are still unknown, there is a growing literature on *why* and *how* schools decided to close and—if they did close—how remote learning affected a range of important outcomes.[1] This growing literature is important, as it provides us a sense of the harm associated with, and the incidence of, remote learning. But it has a critical weakness: the accuracy of the underlying measure of school modality that indicate whether students were receiving remote instruction.

Most of the literature relies on modality trackers, described in more detail below, which classify schools as "in person," "hybrid," or "remote." These trackers, based on self-reports or district websites, do not always agree and often report the same schools or districts as having different modalities. They also categorize schools into distinct modality groups when the reality is often more nuanced. For example, even schools classified as "in person" might offer remote options for certain students, leading to significant variations in the proportion of students who actually attend in person classes. Finally, none of the trackers provide nationwide coverage over time because there is no national database with information about schools' operating status. The

[1] These include, for instance, the factors influencing school instructional modality decisions (e.g., Grossman et al. 2021; Hemphill & Marianno, 2021; Kurmann & Lalé, 2022); the extent to which schooling policy varied across nations (e.g., UNESCO, 2020); whether in-person schooling resulted in COVID spread (e.g., Goldhaber, Imberman et al., 2022); and the impact of school modality on student achievement (e.g., Darling-Aduana et al., 2022; Goldhaber, Kane et al., 2022; Halloran et al., 2021; Kilbride et al., 2021; Kogan & Lavertu, 2021) and behavioral and mental health outcomes (Bacher-Hicks, Goodman, Green and Holt, 2022; Baron et al., 2020; Bauer, 2020; Campbell, 2020; Lee, Ward, Chang, & Downing, 2021; Golberstein, Wen, & Miller, 2020; Hawrilenko, Kroshus, Tandon, & Christakis, 2021; West & Lake, 2021).

bottom line is that while the trackers have been tremendously useful for informing work about the pandemic, they also have important shortcomings.[2]

A second smaller set of studies examines the implications of remote learning cell phone/mobile device usage data to infer when schools are operating in person or not (i.e., identify schools as "remote" when no cell phone pings appear in the school building). The use of these data has intuitive appeal (larger samples; wider coverage, more 'objective'; more nuanced—e.g., degrees of remoteness). These data have been used in studies that examine the correlates of instructional modality and the effects of school modality on various education and health outcomes during the COVID-19 pandemic.[3] Yet here too there are questions about how to translate the volume of mobile device pings inside schools into measures of school modality and questions about how well these measures work for schools in different contexts. To date, only one study that used these data (Parolin & Lee, 2021) has provided evidence of their validity in measuring school modality. But this study uses the data to classify school modality into distinct

---

[2] As Jack and Oster (2023) describe, "Real-time data on school closures was somewhat haphazard during the 2020–2021 school year. Similar to the COVID data, this information was not tracked in a systematic way by any federal agency." Instead, studies have relied on information about school building closures/modality from a single state or from samples of selected schools and districts across the country (see, for example, Parks et al., 2021).

[3] For instance, researchers have used mobility data to show that remote instruction was more common in K-12 schools serving more disadvantaged students and with lower academic achievement, and that there was greater mobility in communities in which K-12 schools operated in person (Courtemanche et al., 2021; Parolin & Lee, 2021). At the postsecondary education level, Huntington-Klein (2020) used mobility data to measure the extent to which colleges and universities were operating in person during the pandemic. Hansen et al. (2022) use mobility data to assess the effect of the pandemic and subsequent school openings on teen suicides, and several researchers are using these data to understand the impact of school openings on economic recovery (Yan, Bayham et al.,2021; Yan et al., 2021; Woody et al., 2020).

categories and does not compare different methodologies for categorization.[4,5] With that in mind, this paper considers how cell phone-based mobility data can be used to gauge whether schools were remote during the pandemic and how they compare to other proxies for remote instruction. Doing so may enable future retrospective studies of outcomes during the pandemic to more accurately measure remote instruction. The findings could also help inform non-pandemic-related studies, for instance those concerning the impacts of temporary school closings due to weather, natural disasters, teachers' strikes, and school breaks.[6]

In this paper we use mobility data obtained from SafeGraph, a location data company, to explore how the data can be used to identify school modalities. We use data from pre-pandemic school years, assuming that schools were operating in person during the school year on non-holiday weekdays, to establish the relationship between mobile device signals (henceforth referred to as "visits") and school modality for schools of varying size and serving different numbers and types of students. Using school modality data collected by the states of Michigan and Washington throughout the 2020-21 school year, we then assess the degree to which the mobility data can accurately predict pandemic-related school building closures. We then apply the models from the Michigan and Washington analysis to both states and nationwide to assess

[4] In their examination of the characteristics of students exposed to remote learning, Parolin and Lee (2021) define schools as being closed for in-person instruction if they experience a 50% or greater year-over-year decline in the number of monthly in-person visits (defined by cellular signals in the SafeGraph data) starting in January of 2019. They also provide several validation checks, such as examining whether the measure of school closure is correlated with survey measures of the share of families reporting distance learning, comparing against school closure survey results collected by Education Week, and comparing against information collected by the authors for the in-person status of two elementary schools per state. All the validation checks suggests that the SafeGraph mobile location data can be used to judge school modality.

[5] Kurmann and Lalé (2022) combine the SafeGraph data with two national school modality trackers – Burbio's national K-12 school modality dataset and the American Enterprise Institute and Davidson College's Return to Learn (R2L) dataset – to construct a measure of "effective in-person learning" during the pandemic in both private and public K-12 schools, but they do not attempt to measure their "effective in-person" metric against an objective measure of school modality.

[6] For instance, see research on summer learning loss (McEachin & Atteberry, 2017); learning disruptions associated with snow days (Goodman, 2014); the implications of the four-day school week (Morton, 2021; Thompson, 2021) and year-round schooling (Graves, 2010; McMullen & Rouse, 2012).

how derived estimates of remote schooling compare to those from two other nationwide sources that are frequently used by researchers (the Return to Learn Tracker and the COVID-19 School Data Hub), and to the method for applying mobility data in Parolin and Lee (2021), henceforth referred to as P&L.

We find that differences between visits to schools in pre-pandemic in-session and out-of-session days are highly predictive of schools being in session. We first generate a predictive logit model of modality during the pre-pandemic period comparing visits on days when schools are in session (non-holiday weekdays) to out of session (weekends) within weeks. We then use Area Under the Receiver Operating Characteristic Curve (AUC) methods to assess the accuracy of our predictions leading to in-sample AUC values over 0.9, where a value of 1 denotes perfect prediction. This suggests the data have excellent discrimination under routine schooling conditions. This finding in and of itself is an important contribution, since it suggests that cell phone data can be used as an inexpensive means to determine when schools are in session and identify the impact of extraordinary events—like snow days or power outages—that keep school buildings closed.

We then apply our methodology to the pandemic school year in Michigan and Washington, comparing our estimates of school modality to state-provided measures. We find somewhat lower discrimination in determining whether schools were in person or remote with the out-of-sample AUC values in the range of 0.70 to 0.80, which is generally considered "acceptable" in the literature on this method (Hosmer, Lemeshow, & Sturdivant, 2013). However, this approach highlights a potential limitation of using mobile phone data and instances where they may not accurately reflect school closures. For example, teachers may still be working in buildings even when schools are closed, and students supposedly learning in person might choose remote learning options instead.

We then use the data to show, in Michigan and Washington, that the accuracy of school modality classification using mobile phone data as measured through type-I (false-positive) and type-2 (false-negative) errors varies significantly depending on the cutoff values used to determine the classification. For example, P&L use a value of 50% drop off in visits from the year before the pandemic as a threshold for classifying schools as remote. As we illustrate, the optimal threshold depends on the weight one places on Type I and II errors and varies by state and school level (elementary, middle, and high schools), suggesting that using a single arbitrary cutoff could hide important information about potential measurement error.

Our paper also illustrates the importance of considering the underlying probability of schools being remote during the pandemic, rather than putting schools into modality categories. For instance, Florida was touted as a state where schools nearly universally returned to in person learning during the start of the 2020-21 school year (Atterbury, 2022; Campo-Flores, 2021). However, reporting also suggests that many families in Florida instead opted for distance learning options (Kam, 2020).[7] The remote schooling data trackers (discussed more extensively below) report that an overwhelming proportion of schools were in person, but the visit data from Florida schools show a much lower proportion of students are likely to have been receiving in person instruction. These kinds of biased measures of school modality during the pandemic could have important implications for understanding the pandemic's—and particularly schools' instructional modalities—impact on outcomes of interest, such as student achievement and

[7] About 35% as reported in Solochek (2020).

mental health. Additional research on determining when and how to use mobility data to accurately measure in-person instruction would enable for better impact analyses in the future.

2. Data

2.1 SafeGraph Data

Fundamental to this study is the use of SafeGraph's mobile Global Positioning System (GPS) data. SafeGraph is a data company specializing in location and points of interest data. The company provided the data to us through a public access agreement. The data include anonymized GPS records from mobile devices, which enables users of the dataset to assess social distancing during the pandemic, full-time work, and foot traffic—including foot traffic in schools.[8]

SafeGraph collects and maintains aggregated anonymized location data from approximately 45 million mobile devices in the US, or roughly 10% of all US mobile devices.[9] These data allow us to study US population mobility patterns before and during the COVID-19 pandemic. Specifically, we received information on the number of unique mobile device cellular signals to schools each day from the beginning of September 2018 through the end of June 2021, covering one full pre-pandemic year (i.e., the 2019-20 school year, during which nearly all school buildings closed in March), and the 2020-21 school year, which saw variation in school modality. As previously stated, we refer to these cellular signals as "visits."

[8] SafeGraph data are recorded for all mobile devices, including cell phones, tablets, personal navigation devices, smart watches, etc., unless the individual using the device is known to be below age 16, an issue we return to below. The data record the unique visits made by mobile devices utilizing mobile apps with a GPS component. When a device is observed within the boundary of a school it is recorded as having visited that school. Because cell phones represent the lion's share of mobile devices, we use the terms "mobile devices" and "cell phones" interchangeably throughout the remainder of the paper.

[9] SafeGraph's collection and distribution of foot traffic data was discontinued in 2023, but this data was widely available to researchers before that time. This study serves to evaluate the value of existing stores of the SafeGraph data for estimating school closures, and more generally the efficacy of using this type of mobility data.

SafeGraph data have been used to track stay-at-home orders (Yan, Bayham et al., 2021), visits to treatment facilities and medical centers (Cantor et al., 2021), and other issues related to mobility during the pandemic. It is important to note, however, that there are limitations when using the SafeGraph data. First, the data do not include information from devices without service, devices that lack an application with tracking services, and/or from users who have opted out of location services. Second, and particularly relevant to our focus on schools, SafeGraph states that data on users aged 16 or younger are not collected (SafeGraph, n.d.-b), meaning that estimates of school instructional modality will be based on students 17 years of age and older, and adults (e.g., teachers, counselors, administrators) in the buildings.[10]

There are other reasons why the SafeGraph data require careful interpretation. For instance, the presence of some mobile device visits at schools during the pandemic does not necessarily indicate that schools were open for in person instruction. Likewise, the absence (or reduction) in visits does not necessarily indicate that schools were not open for in person instruction. As has been reported (Cohen, 2020), teachers may have been delivering instruction remotely to students from inside school buildings. This type of activity would show up as mobile device visits, so long as those entering schools carried tracked devices. Similarly, the absence of significant numbers of mobile device visits does not necessarily indicate schools were not operating with traditional in person instruction. For instance, if significant shares of students remained learning remotely even when their districts were offering in person or hybrid options, or if social distancing rules reduced the numbers of adults present at the building —i.e., parents

[10] SafeGraph determines whether a user is above or below the age cutoff by using birthdate information that is entered into various mobile apps.

and other non-staff visitors—schools designated as in person could look considerably different from locale to locale.[11]

Finally, elementary and smaller schools are likely to have fewer individuals carrying cell phones or to have SafeGraph data attributed to under-16 mobile device users, for whom data is not collected. In addition, there may be systematic differences in the number of mobile devices recorded at schools based on differences in rates of cell phone ownership, which varies across demographics like race/ethnicity and poverty status (Pew Research Center, 2021). Thus, there may be fewer recorded visits from mobile devices in schools in communities with higher proportions of low-income students and teachers and students of color.[12] All of these are reasons to assess the accuracy of visits for different types of schools (described below).

2.2 Data Linkages and National Sample

To link SafeGraph data to individual schools we first restrict the data to only include points of interest (POIs) assigned school-related codes.[13] We use probabilistic matching to link the SafeGraph data from school-related POI codes to K12 public schools via data maintained by the National Center for Education Statistics (NCES). Specifically, we use NCES data from the 2018-19 Common Core of Data (CCD) and the US Census GIS Education Demographic and Geographic Estimates Program (EDGE) data. These data contain information on the physical location of every public school in the United States, a phone number and address, and latitude

---

[11] For a descriptive picture of how school modality varied in the 2020-21 school year, see Sachs et al. (2022). Hopkins, Strunk, and Kilbride (2022) also examine the variation in school modality and in student uptake of in-person instruction in Michigan over the course of the 2020-21 school year.

[12] Comparison of SafeGraph sampling rates at the Census Block Group level against Census data shows that there is geographic bias in the sample, but this bias does not lead to considerable over- or under-sampling by race, income level, or education level (Squire, 2019). However, this addresses over-sampling in the aggregate, not among schools specifically, and data in under-sampled geographies should be noisier, making inferences about school modality more difficult in areas with less mobile device data.

[13] A point of interest is a set coordinate point tied to a particular place or location of interest. In this paper we only consider POIs that are flagged as schools and match to the Common Core of Data.

and longitude coordinates. The sample includes schools appearing in both pre-pandemic (2018-19) and pandemic years (2019-20 and 2020-21).[14] We match just over 84,000 POIs in the SafeGraph data to public schools in the federal data. The CCD includes information about key school characteristics, such as the racial makeup of students and the number of students who receive Free or Reduced Priced Lunch (henceforth FRL, a proxy for poverty status). The CCD data also contains a flag for different school types (Regular, Alternative, Career and Technical, and Special Education only schools). To avoid instances where non-traditional schools might exhibit divergent or odd patterns in enrollment, we restrict the sample to include only Regular schools. This results in a final sample of 78,905 or just over 85% of all public regular schools.

For our analyses, we aggregate the SafeGraph data to the school-by-week level by taking the average number of visits originating in each school during the week in which schools are in session. *Each school has two observations per week: one for in-session days and one for out-of-session/closed days.* This aggregation serves several purposes. First, it reduces noise in the individual observations by averaging visitor counts over multiple days. SafeGraph data can be a fairly noisy measure from day to day, but as suggested by SafeGraph documentation (SafeGraph, n.d.-a) this can be averaged out over broader time periods. This should reduce overfitting on the day-to-day variation that is not of interest in the model. Second, in most weeks, the ratio of open to closed days is five to two. In classification, the prediction for any given day will over-favor predictions that the school is open. Aggregating the data to have an equal number of open and

[14] While this restriction does eliminate some schools, over 97% of public schools appear in all three years. The EDGE and CCD data contain indicators such as phone number and address, as well as latitude and longitude coordinates which we use to perform probabilistic matching to the SafeGraph data. We run a series of probabilistic matches, first using phone number, then address and name. Following the potential match, we analyze if the latitude and longitudes for each NCES school and SafeGraph POI are within 200 meters.

closed observations each week avoids this issue.[15] Third, aggregation considerably reduces the computational power necessary to estimate the models.

For the pre-pandemic analysis (described below in Section 3), we use data from MDR, an education data company that provides school and district calendar information related to school year start and end dates, as well as major breaks throughout the year (Winter, Mid-winter, and Spring breaks).[16] These data are consistent with the data we described above in that they cover the 2018-19, 2019-20, and 2020-21 school years and represent 98,626 unique schools. The MDR data lack information on some types of school closures, such as snow and teacher in-service days, but otherwise provide a clear indication of whether districts had students attending schools in person on a given day or week. While there are different ways to account for holidays, during the pre-and post-pandemic periods we remove visit counts during known holidays because these are days schools are officially out of session (there are also times during holidays when school buildings are used for events or gatherings, which we want to avoid treating as visits).

Figure 1 shows the average number of school visits per week (includes both in- and out-of-sessions days) in the 2019-20 (blue line) and 2020-21 (yellow line) school years, and the gaps between them (shaded blue or yellow). We compare both the visits from the first week of September through the last week of June for the nation (top panel) and for the two focus states in the study: Michigan (middle panel) and Washington (bottom panel). The nationwide closure of schools that occurred roughly in the third week of March 2020 (Ujifusa, 2020) is marked with a vertical red line.[17]

[15] Alternatively, this could be achieved by weighting each open/closed day by the inverse of the number of open/closed days in that week.
[16] For more on the MDR school calendar data see: https://mdreducation.com/public-school-open-and-close-dates/
[17] For later analyses we set the pandemic period for all weeks beyond March 9, 2020. Michigan and Washington issued their school closure dates on March 12 and March 17 respectively.

The effect of the pandemic and vacations on visits is readily apparent from the figure. For instance, visits drop precipitously from March to April of the 2019-20 school year to roughly the level in the late December/early January period, when schools are typically in recess. The orange shaded area of the figure depicts the difference between the early and late pandemic years and shows that visits were consistently higher in April through June 2021, when schools largely re-opened for in person instruction, compared to the same period in 2020. The figure also shows large dips in visits during periods when schools are out of session, during June and holiday breaks.[18]

In Panel A of Table 1 we present summary statistics comparing the characteristics of schools matched and unmatched to the SafeGraph analytic sample. Schools in our matched sample are larger, have lower proportions of under-represented minority students (i.e., Black and Hispanic students), are higher-performing, and are more likely to be located in suburbs and less likely to be located in rural or town areas.

In Panel B of the table, we compare the average number of visits in the 1st week of October 2019 to October 2020 between the pre- and post-pandemic school years for different kinds of schools. Consistent with Figure 1, there are far fewer visits in the fall of 2020 relative to the fall of 2019 across the board, but particularly in larger and lower-income schools, middle and high schools, and in schools in urban and suburban areas. The difference is particularly large when comparing the schools with the highest and lowest share of underrepresented minority students (URMs), with the former having far larger drops in foot traffic. These basic descriptive findings support earlier studies using other datasets that suggest urban, larger, and lower-income

[18] Reviewing the MDR school calendar data we see that approximately 93% of schools were recorded as on break during the 3rd week of June, and 92% during the last week of December.

schools were less likely to be open for in person instruction during the 2020-21 school year (e.g., Grossman et al., 2021; Hopkins, Strunk, & Kilbride, 2022).

2.3 Michigan, Washington, and National Tracker Data on School Modality

To assess the accuracy of SafeGraph visit data in predicting school modality during the pandemic, we also link the visit data to state-level data on school modality from Michigan and Washington. These modality data are collected by each states' respective state education agencies [the Michigan Department of Education (MDE) and Michigan's Center for Educational Performance Information (CEPI) and Washington's Office of the Superintendent of Public Instruction (OSPI)]. While we cannot directly assess the accuracy of the state-level school modality data, it was collected for what might be considered high stakes purposes, as each of the state education agencies were making school reopening decisions at the time of data collection and there was also intensive media coverage about school modality at this time (e.g., Furfaro, Villa, & Bazzaz, 2021; Eggert, 2021).

The data are relatively consistent across both states, though there are a few differences. In Michigan, districts were surveyed about how they planned to deliver instruction in each upcoming month. In Washington, districts reported the modality that was offered during the final day of the month from Fall 2020 through the end of the calendar year and then the planned modality for the upcoming week beginning January 18, 2021. We compare Michigan districts' modalities at the beginning of October to Washington districts' modalities beginning September 30 and the surveys for the last week of the previous month for February–April 2021 for both states.

The definitions of the instructional modality categories (i.e., in person, hybrid, and remote) vary slightly between the two states. In Michigan, the definitions are based on what districts provided their general education students. In person districts are defined as those that provide general education students with the option of full-time in person instruction.[19] If general education students are in person a portion of the week—usually two to three days or parts of each day—but are not in person full-time, then the district is recorded as "hybrid." If districts provide all instruction in a remote or virtual format for their general education population, then they are assigned "remote."[20] In Washington, "in person" districts are those that provide "typical/traditional in person" instruction to elementary, middle, and/or high school students; "hybrid" districts are those where all students received "partially in person" instruction or the district used a "phase-in" approach (i.e., some students received partially or fully in person instruction while others received remote instruction); and "remote" districts are those in which all students, or all except small subgroups of students, received fully remote instruction.

Table 2 examines patterns of school characteristics across different instructional settings in Michigan and Washington. Using the modality for the first week of October 2020, we compare the differences in mean characteristics between schools operating in person, hybrid, and remote. We note that these data were collected at the district level, and we assign the recorded modality to each school in the district. This means that if there were districts that offered different modalities by school level, we would not capture them. In these descriptive results (and those below), we include the hybrid category. But as we describe below, there is considerable ambiguity surrounding the definition of hybrid, so we opted to eliminate it in the estimation of the regressions described in Section 3. Eliminating hybrid results in a decrease of roughly 45% of observations nationally, and 42% and 55% fewer observations in Michigan and Washington, respectively.

[19] Students may opt for either hybrid or remote instruction but the majority of students are in the in-person option.
[20] In Michigan the categorizations do not take into account the modality of instruction for special education students.

There are three differences between the distribution of modalities across the two states. First, the great majority (83%) of schools in Washington were operating remotely in October of 2020, whereas in Michigan only 16% of schools were operating remotely and the majority—62%—were offering in person instruction. Second, whereas in Washington remote districts were substantially larger, on average, than hybrid or in person districts, in Michigan the differences were less stark, and hybrid districts were larger, on average. Third, in Michigan, remote districts had substantially higher proportions of students who qualified for free- and reduced-price lunch and hybrid districts had relatively low proportions of low-income students. By contrast, in Washington, districts of all three modalities had approximately the same proportions of low-income students. We see similar trends in modality distribution across geographic types. Urban districts were far more likely to be operating remotely in both Michigan and Washington, whereas town and rural districts were more likely to be offering in person instruction. In addition, high schools in both states were more likely to be operating in person.[21]

We also compare estimates derived from the SafeGraph data during the 2020-21 pandemic year to two national sources of data about school reopening being used by researchers: the Return to Learn Tracker (R2L) and the COVID-19 School Data Hub. We use these sources of information because they cover a larger number of districts over a greater time span than either the Michigan or Washington data.[22,23] Both datasets are created from a variety of methods including website scraping, direct collection, crowdsourcing, or a combination of these methods.

[21] For more details about the Michigan and Washington data collection and changes over time in school modality in those states, see Goldhaber et al. (2022).

[22] Examples of other modality trackers over the pandemic are Burbio (see: https://cai.burbio.com/school-opening-tracker/); MCH (see: https://www.mchdata.com/covid19/schoolclosings); and Education Week (see: https://www.edweek.org/leadership/map-where-are-schools-closed/2020/07).

[23] Generally, the coverage of school districts represented in these alternative sources is less than in the RL2 and Data Hub datasets. For instance, the Edweek, Burbio, and MCH data cover 1,200; 900; and 10,500 school districts, respectively. These datasets also vary in the number of weeks of data collected and the frequency with which the data were updated. In some cases, for instance, the MCH data were only collected once per semester. To our

The R2L[24] dataset, "monitor[ed] roughly 8,500 public school districts' instructional statuses on a weekly basis" by compiling district reopening information from a combination of external data (e.g., data from MCH Strategic Data; see footnote 23) and those collected by web scraping. These data provide information on reopening and instructional status from August 2020 through June 2021 (i.e., 45 weeks of modality data). During the period we focus on, the R2L data contained observations from all 50 states and the District of Columbia.

The COVID-19 School Data Hub, School Learning Model Database[25] is derived from data requests submitted to state education agencies (SEAs) for information from schools or districts, at the most frequent time periods available. The data span from June 2020 through the end of June 2021, however the bulk of collection did not begin until August 2020. Data Hub modality data cover roughly 10,000 unique school districts (representing 63,834 schools) and had consistent observations for 35 states.[26] We merged the two trackers when they had consistent weekly observations of modality. This resulted in a tracker dataset representing 5,586 unique districts and 43,886 unique schools.

3. Analytic Approach

We use four different modeling strategies to assess how well SafeGraph data measure school closures. We then assess the accuracy of various school modality databases used in the literature from our SafeGraph-based predictions and describe the distribution of school modalities across students (along the lines of P&L, 2021). For all models we estimate (at the

knowledge, there is scant evidence demonstrating the degree to which modality data overlaps from source to source and the relative accuracy of these data sources more generally.

[24] A school reopening database developed by the American Enterprise Institute in partnership with the College Crisis Initiative at Davidson College, for more see: https://www.returntolearntracker.net/

[25] More information on the database can be found at: https://www.covidschooldatahub.com/about

[26] There are some additional states (e.g., Florida, Texas, Pennsylvania, etc.) represented in the Data Hub data however, collection took place at only two points during the school year and hence these observations were dropped due to infrequent collection. Note also that these figures are based on schools represented at the time of writing, and do not necessarily reflect any subsequent updates made to the respective modality tracker datasets.

school-week-session level) the likelihood that we are observing a day in which schools are *in session* and *in person*.[27] The assumption is that all the schools that are in session pre-pandemic are also in person, but this varies during the 2020-21 school year.

3.1 Estimating the Extent to Which Pre-Pandemic SafeGraph Visit Data Accurately Predict Schools Being In Session

Figure 1 indicates that the SafeGraph foot traffic data are informative for determining whether a school is or is not in session. However, we want to understand the accuracy of these data in predicting school closures. To do so, we first examine the pre-COVID data to understand how visits differ on break days and in-session days and how this varies by school characteristics. We create the variable $InSession_{jwk}$, which is an indicator for whether school i in district j is scheduled to hold classes on day d in week w, using MDR calendar data. To reduce noise, we then aggregate the data to two observations ($k \in [0,1]$) per school per week: one observation for out-of-session days and one for in-session days. In regressions, the observations are all weighted by the number of days used in the aggregation (e.g. weeks with 4 in-session days and 2 out-of-session days have weights for each observation of 4 and 2, respectively).[28] We use this to gauge how well the model predicts openings outside of a pandemic context by estimating logit models of the form:

$$\Pr\left(InSession_{jwk}\right) = logit\left(\beta_0 + \beta_1 \operatorname{asinh}\left(Visits_{ijwk}\right) + \boldsymbol{X_{ijw}\Omega_1} + \mathbf{S_w}\boldsymbol{\beta_2} + \operatorname{asinh}\left(Visits_{ijwk}\right) \times \boldsymbol{X_{ijw}\Omega_2} + \varepsilon_{ijwk}\right) \quad (1)$$

[27] Out-of-session days are weekends. We drop school holidays from the sample. The aggregation to the school week level means that each school has two observations in each week. One of these observations includes the average number of visits on days when the school is in session. The other observation includes only weekends. Thus, in a typical school week there will be 5 in-session days, but that will vary depending on school holidays.

[28] Weeks where there are no in-session days, e.g., during school system summer vacations, are eliminated from the sample.

using the MDR and SafeGraph data where *Visits* are the average daily number of devices observed at the school in the in-session or break days, respectively, during week w,[29] and **X** is a series of school characteristics including quartile indicators for the percent of students from an underrepresented minority, the percent of students receiving free or reduced-price lunch, and total enrollment, as well as urbanicity. Since *Visits* tend to be distributed with a long tail, we use an inverse hyperbolic sine transformation ("asinh", or $\text{asinh}(x) = \ln(x + \sqrt{x^2 + 1})$) to better reflect the structure of the data.[30] S_w is a seasonality control: the average visits across all schools in that same calendar month in 2018. The controls account for the potential of differences in mobile phone usage across demographics and the number of devices inside the school.[31] In particular, the interactions between the controls and *Visits* allow for the explanatory power of *Visits* to differ across school characteristics. Allowing for these interactions is important as schools vary considerably in terms of the number of individuals with trackable mobile phones, thus an additional visit can have very different marginal impacts on the predictions.

We estimate Equation (1) focusing on data from September 2018 through February 2020 (i.e., the period prior to pandemic onset when we have the nationwide MDR school calendar data) to estimate how many visits a school receives on a typical school day relative to non-school days separately for elementary, middle, and high schools. This allows us to provide a baseline for how much we might expect visits to fall during a typical day when schools are closed but when we know there is no remote instruction occurring. As described above, we look within

[29] In all applications in this paper, visits are adjusted for the SafeGraph sample size, which changes over time. Within each state, raw visit counts are divided by the sample size in that state in that week, and then multiplied by the all-time state-level average of sample size.

[30] We use the inverse hyperbolic sine instead of the more standard logarithmic model because of the large mass of zero values for *Visits*. Note that we are concerned with accounting for skew and not about maintaining a percentage-increase interpretation, which makes the IHS transformation appropriate.

[31] While the SafeGraph data are not supposed to include individuals under 16 years old, this only applies if the person with the mobile device provides their age in an app that shares data.

each school and week to calculate differences in trafficked days (scheduled school days) and less trafficked days (weekends and scheduled school holidays).

3.2 Validating the Predictive Model

Next, we take the prediction models for Michigan and Washington and evaluate their accuracy using an Area Under the Receiver Operating Characteristic Curve (AUC) procedure. The AUC is a measure of how well a model with a binary dependent variable predicts outcomes. Specifically, we use the models in Equiation (1), estimated using pre-pandemic data to classify a school as in person or remote during the pandemic (for the time being, we drop hybrid schools as they are harder to identify through this method) by setting a threshold where a predicted probability falling above indicates successes (schools predicted to be in person), and all observations with predicted probabilities below the threshold as failures (schools predicted to be remote). The Receiver Operating Characteristic (ROC) curve considers each possible choice of threshold, varying from the lowest possible value to the highest, and then maps the resulting false positive rate along the x-axis, and the true positive rate along the y-axis.

The AUC is the area underneath this curve, which in effect is the average true positive rate, across all possible values of the false positive rate. Thus, an AUC = 1 indicates that the model is perfectly positively predictive, an AUC = 0 indicates it is perfectly negatively predictive, and an AUC = 0.5 indicates the model essentially picks success and failure at random. Typically, an AUC above .7 indicates the model is acceptably predictive, and .8 is excellent (Hosmer, Lemeshow, & Sturdivant, 2013), although there is no particular meaning to the .7 and .8 cutoffs. In general, a higher value indicates better classification ability.

4. Results

We organize the discussion of the results as follows: In 4.1 we provide descriptive evidence about the agreement/disagreement nationally and in the focus states among several of the major modality trackers. In 4.2 we describe the extent to which SafeGraph data can be used to characterize schools being in session or remote in a *pre-pandemic period*. This is important, since knowing at scale whether schools are open is valuable for research on issues such as the impact of snow days and time spent in school (e.g., Goodman, 2014; Marcotte & Hemelt, 2008) or different school calendars (e.g., McMullen & Rouse, 2012; Morton, 2021). We show the accuracy of different methods of estimating whether schools were in person or remote during the 2020-21 *pandemic* school year in Section 4.3. Finally, in Section 4.4 we compare what the various measures of school modality in the 2020-21 school year suggest about the prevalence and incidence of remote schooling.

4.1 Modality Tracker Agreement/Disagreement

Both school modality trackers are somewhat geographically limited. For instance, more than 20% of rural schools are not represented in either the R2L or Data Hub data. Even when the two trackers had data from the same schools, they frequently categorize instructional modality differently. We illustrate this in Table 3, which displays the percent of schools observed in each instructional modality (in-person, hybrid, or remote) recorded in the R2L data versus in the Data Hub, for the national sample (Panel A) and our two focus states (Michigan in Panel B and Washington in Panel C). The denominator used to determine the percentages in each cell is the row denominator, so the columns do not consistently add up to 100.[32]

Focusing first on the national sample, we see that while most of the modality measures are aligned between the two datasets, there is less agreement than one might expect. For instance, just over 85% of schools coded as in person in the R2L data are similarly coded in the Data Hub data (shown in the upper left cell of Table 3). There is similar agreement in the remote

[32] The rows will sum to hundred as the denominator is based on the total for the R2L data, consequently columns (based on Data Hub data) may not sum to 100.

instruction category, but in both cases, there are still approximately 15-20% of schools classified differently between the two sources for in person and remote. Hybrid modalities appear the noisiest across the two datasets.

Agreement for the hybrid category is similarly low in the two state samples, though the agreement for remote is also low (less than 70%) in Michigan. The low agreement for the hybrid category is not surprising given that the definition for this category varies between data sources and across states and districts. For example, the Data Hub defines hybrid as "A blend or combination of in person and virtual instruction for all or the majority of students." R2L defines it as "Either students in some grades can return to buildings in person while other grades can only return in a hybrid or remote model or all students can return to buildings for four days or less each week (or five partial days) while learning remotely from home the remaining time."[33] There are also differences across the focus states in terms of how the *states* define hybrid instruction.

4.2 Predicting Pre-Pandemic In-Session Schools

Our focus is on the degree to which the adjusted count of visits to schools is predictive of school modality and whether this is sensitive to enrollment counts or the demographics of students in schools. Our preferred specification, because of goodness of fit, utilizes the adjusted count on in-session and out-of-session days during the school year.[34] Chow tests confirm that the fit is better for models estimated separately for elementary, middle, and high schools, rather than combining across school types and including type indicators.[35]

---

[33] The definitions for in person and remote are more closely aligned between the two data sources.
[34] Other approaches that performed less well included: (a) the ratio of adjusted counts on in-session days to out-of-session days, (b) the difference of adjusted counts on in-session days to out-of-session days, and (c) count relative to enrollment.
[35] Chow tests strongly reject the null ($p < .000001$) that the coefficients are the same when we instead estimate a stacked model across school types.

Table 4 reports the logit coefficient estimates at the national level from various specifications of Equation (1) for elementary, middle, and high schools in columns 1-3, respectively. Putting one of these effects in context, the .874 coefficient on the adjusted visit count for high schools means that a one-unit increase in $\text{asinh}(Visits)$ implies an increase in the probability of in person schooling rising from 40% to 61.9% among schools in the first quartile of enrollment, FRL, and URM. A one-unit increase in $\text{asinh}\ (Visits)$ is equivalent to an increase in visit counts from 20 recorded visits, which is the overall mean immediately following the pandemic, to 54.4 recorded visits ($\text{asinh}(20) = 3.69, \text{asinh}(54.4) = 4.69$). This prediction controls for the quartiles for school enrollment, percent of underrepresented minority students, URM, and FRL students, which vary in some ways between elementary, middle, and high school, as well as across interactions between these and visits.[36]

Figure 2 shows the receiver operating characteristic (ROC) curves for the above models in their ability to classify pre-pandemic data into in-session vs. out-of-session days; the accompanying AUCs for elementary, middle, and high schools are: .940, .941, and .922, respectively.[37] While there is no hard and fast standard for judging the predictive capacity of the model, Hosmer and Lemeshow (2000) advance a rule of thumb suggesting that these models are highly predictive (AUC values of 0.7-0.8 suggest acceptable discrimination, 0.8-0.9 are excellent

[36] There is no conceptual reason to believe that the main effect of school covariates should be predictive of school modality, because the structure of the data artificially reinforces that open and closed days are balanced within school. Significant findings on the un-interacted covariates are an example of collider bias (Elwert and Winship, 2014) which arises because we are including visit count, which caused both by modality and school characteristics. Adjusting for count induces a relationship between characteristics and modality. If we omit visit counts as a predictor, all school characteristics have insignificant near-zero coefficients.

[37] The ROC curve plots the fraction of observed positive outcomes (schools being open) that are correctly classified (the “sensitivity”) against [1 – the “specificity”], which is the fraction of negative outcomes (schools being closed) that are correctly classified for different cut points (measured along x-axis) for classifying the predictions. The ROC curve runs from 0 to 1 with the x-axis representing the values of 1-specificity (the false positive rate) and the y-axis measures sensitivity (the true positive rate). A model with no predictive power (beyond chance) would have a ROC along the 45-degree line and greater predictive power is represented by ROCs that are toward the upper left quadrant of the figure.

discrimination, and values greater than 0.9 show outstanding discrimination). The model appears to be effective at using visit counts to classify observations in the pre-pandemic period as either being in session or out of session It's important to note that this prediction uses data from the same time period it is analyzing. This approach generally gives better results than using data from before the pandemic to make predictions about what happened after the pandemic, as we do below.

The inverse hyperbolic sine specification and interaction terms make the magnitudes of each of the coefficients difficult to interpret. However, we attempt to make the relationships between visit volume and schooling characteristics clearer in Figure 3. Each panel (Figure 3a – 3i) shows how the predicted likelihood of schools being open (on the y-axis) changes as the adjusted visit count deciles (on the x-axis) increase.[38] Each line in a panel shows the relationship for a given quartile of a demographic characteristic (enrollment, URM share, and FRL). We do this separately for elementary, middle, and high schools. For instance: Figure 3a is the predicted probability of elementary schools being in session for the four quartiles of enrollment (holding URM and FRL constant) at each decile of the adjusted visit count.

As we can see from all the estimates, the probability of schools being in session sharply increases with the decile of visit count. We also see that there is relatively little change in the estimates of schools being in session based on the quartiles of free-or reduced-price lunch or underrepresented minority students; the spread between the quartiles is small (in rows 2 and 3). However, the quartile of enrollment does appear to matter considerably (see row 1). For instance, when looking at enrollment in middle schools, the predicted probability of being in session for the lowest quartile is .765 at the median level of visits versus a predicted probability of .444 for

[38] The deciles are based on national distributions from average visits in a week within school type (elementary, middle, and high schools).

the highest quartile, a differential of .321. For comparison, the differential between the predicted probabilities for high- and low-quartile FRL middle schools is only .087 and high- and low-quartile URM middle schools is only .166.

4.3 Predicting 2020-21 School Modality in Michigan and Washington

We now turn to using the above model to predict school modality in the two focus states during the pandemic year of 2020-21 to determine how well our out-of-sample predictions apply to pandemic related modality. Specifically, we use the estimates reported in Table 4 to generate predictions of school modality in 2020-21 in Michigan and Washington, and compare these to estimates taken from a P&L-like approach in which a range of reduced-traffic cutoffs are considered.[39] AUCs using the P&L approach are very slightly larger than ours at each school level (see Table 5) and for both states.[40] Note, however, that the P&L method cannot be used in a regular school year since there is unlikely to be a significant year-to-year change in visits outside of major events like the COVID-19 pandemic.

To further compare our methodology to P&L, in Table 6 we apply several different probability thresholds to classify schools as in person or remote and assess the proportion of times that each approach correctly predicts the in person and remote school modalities (e.g., the rates of true positives and negatives assuming that the data on in person and remote schooling collected by each state are in fact correct). Recall that P&L use a cutoff of 50% reduction in visits from the year before the pandemic to the pandemic year to classify schools as being

[39] We also attempted a variant of this approach that used within-school changes in visits from pre-pandemic to post-pandemic, in which the weekday declines relative to weekend declines were a measure of schools shifting away from in-person instruction. However, this approach suffered from poor predictive ability because weekend visits declined by proportionally similar amounts to weekday visits, indicating that remote schools also reduced their (already low) weekend foot traffic activity, making this untenable as a way of classifying remote schooling.
[40] We use the out of sample estimates for Michigan and Washington state. The in-sample (i.e., during the 2021-22 school year) estimates of school modality have AUC values that are about .07 higher for each school level when using post-pandemic data to predict post-pandemic outcomes, or .2 higher when using pre-pandemic data to predict pre-pandemic outcomes.

remote. Since there is a fundamental tradeoff in accuracy of the model correctly identifying remote modality versus correctly identifying in person modality, we adjust the threshold to focus more on accurate prediction of one modality or the other. A weight of 50% is equally concerned with prediction of both modalities, while a remote weight of 75%, for example, places three times more weight on correctly classifying each case of actual remote schooling correctly than on correctly classifying each case of actual in-person correctly.

We can gain several useful insights from Table 6. First, there is no single best cutoff to use. As is inherent in any classification problem, changing the cutoff makes it easier to correctly classify one category but harder to correctly classify the other, so the best cutoff depends on whether it is more important to correctly spot remote days or to correctly spot in person days. Second, comparing our method to P&L, there is no clear winner—the best approach depends on whether the user places higher value on correctly identifying remote schooling or in-person schooling. For most school level/weight combinations, one method is better at classifying remote and the other is better at classifying in person. In general, it appears that our method tends to do best when there is more weight on remote classifications, and P&L's method does best when there is more weight on in person classifications, but not in all cases.

The specific thresholds chosen are also of interest. The thresholds for P&L show that the original choice of a 50% drop is roughly close to the optimal threshold if it is three times as important to identify remote days as in person days for middle and high schools. However, for elementary schools, a 50% drop treats them as equally important, and if properly classifying remote days is more important, then smaller drops should also be classified as remote. Note, however, that the optimal thresholds differ to some degree across states and school levels. For instance, equally weighting correct classification of in person and remote middle schools is estimated to have a visit drop-off cutoff of 61% in Michigan and 77% in Washington.

For our method, optimal cutoffs are very high, generally above 0.8, unless one has very high weight on identifying in person modality correctly. One might expect that, since these cutoffs are on a probability scale, ideal classification might occur around a cutoff of 0.5. However, the probability of a given pre-pandemic day being remote changes sharply over the very low range of visit counts, which, in turn, cover a very large portion of the hyperbolic sine-scaled visit count range. This wide range of scaled data covering the remote days pushes the effective classification cutoffs for classifying in person days far up the probability scale, and so optimal cutoffs are generally much higher than 0.5.

4.4 Estimates of the Prevalence and Incidence of Remote Schooling in 2020-21

There are several studies that report differential incidence of school modality across student demographics, often with greater incidence of remote schooling for schools serving higher proportions of Black and Hispanic students (e.g., Camp et al., 2023; Parolin & Lee, 2021; Grossmann et al., 2021). This is consequential, as remote schooling has been found to be an important contributor to the disproportionate impact of the pandemic on the test achievement of URM students (Goldhaber et al., 2023).

In this section we revisit this issue using different methods to derive estimates of the incidents of remote schooling. We replicate P&L's estimates of the prevalence and incidents (across different student demographics) of remote schooling during the onset of the pandemic and subsequent school year in Figure 4. In the right panel of Figure 4 we recreate what the share would look like using individual school probabilities of remote schooling. We do this by taking a probability-weighted average of the likelihood of remote schooling across all schools at the school-week level. Letting *d* represent a student demographic group (Asian, Black, Hispanic, White) and *t* represent the week, the estimated proportion remote is:

$$Remote_{dt} = \sum_j Prob_j * Share_{dj} \quad (2)$$

In Equation (2) above, $Prob_j$ is the probability that school j is remote (calculated from the findings in Table 4) at time t and $Share_{dj}$ is school j's proportion (relative to the nation) of students in racial/ethnic demographic d.[41]

There are differences between our sample of schools and the level of data aggregation between what we do and P&L's work.[42] Despite these differences, we closely replicate P&L's findings in the left panel of Figure 4.[43] But, as is readily apparent from visual inspection of the right panel of Figure 4, we get quite different estimates of the prevalence and incidents of remote schooling using our methodology. For instance, we estimate that less than 70% of students in any demographic were in remote schooling in April 2020 compared to P&L's estimates that suggest over 90% of students were in remote schooling in the same period. And while, as noted above, there is no comprehensive national data on school modality, there are good reasons to believe that our estimates are less likely to be correct, at least during the early stages of the pandemic. As was popularly reported in the press (e.g., Peele & Riser-Kositsky, 2020), effectively all schools were closed at the end of the 2019-20 school year (hence the P&L estimates seem much more aligned with what one might expect).

We showed in Section 4.2 that our approach to estimating schools being in session in a pre-pandemic period performs reasonably well. The main problem with our approach in

[41] We used school enrollment data from the Common Core of Data to calculate these shares.
[42] P&L use district-month data whereas we aggregate to the school-week level.
[43] See Figure 2 on page 3 in P&L (2020). We further verify this by aggregating our weekly data to the month level (the level used by P&L) and correlating the share of students of demographic D in each month from P&L with our estimates. The correlation across demographics and months where data overlap (March 2020 to December 2020) are over 0.7. We thank Zach Parolin for providing the data that allowed for this correlation.

estimating the onset of the pandemic-related school closures is that our model is sensitive to changes between in-session and out-of-session visits at the low end of the scale, and the scale of visits to schools change from pre-pandemic to pandemic by a magnitude that is mostly outside the range of what we estimate using an in-session/out-of-session contrast. Put another way, the magnitude of difference between in-session and out-of-session visits before pandemic is dwarfed by the drop off from the pre-pandemic to pandemic period in visits, which we do not capture in our model, but the P&L approach does capture.

The picture in the following school year appears more complicated. Consider, for instance, that most K-12 schools in Florida reopened for in person learning at the beginning of the 2020-21 school year (Doyle et al., 2021). We thus might expect the various means of estimating remote schooling would tend to agree. However, they do not. In August and September of 2020, the R2L national modality tracker suggests 25% of students were in remote schooling; the P&L method suggests that approximately 50% of students were in remote schooling; and our probability method suggests it was about 75%.[44]

We conduct several additional comparisons between estimates of the share of students in remote schooling. Specifically, we compare estimates derived from our methods described above (using optimal thresholds reported in Table 6) to those used by P&L (using the 50% drop-off in visits as the threshold) and those from the R2L national modality tracker. We do this for the 2020-21 school year in Michigan and Washington, where we can juxtapose each of these against the information collected by the states and the nation data.

We calculate the share of students who are in remote schooling for elementary, middle, and high schools in each state as follows. For the share according to the state or R2L data (where

[44] DataHub data are not considered in this comparison because of infrequent sampling.

schools are designated as remote or in person), we simply take the number of students in schools classified as remote at each school level (elementary, middle, and high) and divide by the total number of students in the district at that level. For the share according to a method that relies on a threshold for classifying whether schools are remote or in person or a threshold value (e.g., the P&L 50% drop-off or the optimal values reported in Table 6), we categorize schools as remote or in person, and then take the number of students in schools categorized as remote divided by the total number of students in the district. Finally, the probability method uses the estimated coefficients from Table 4 applied to the characteristics of each school in each district. We then estimate the district percentage as the weighted average across schools in the district at each level (where the weights are the share of students at the school relative to the district).

We report correlation matrices in Table 7 showing how the various estimates compare in Michigan (Panel A) and Washington (Panel B) by school level. These comparisons are based on aggregations to the district-month-school in each state.[45] In Michigan, at the elementary level, there is relatively little difference in the correlations between the various methods to calculate remote share and the remote share based on the state data (the correlations range from 0.48 to 0.54). The results are similar in Washington, where the correlations range from 0.56 to 0.70. But in both states the correlations from the various means of estimating remote status and the state data are far higher than the correlation between the R2L collection and the state data (0.11 in Michigan and 0.18 in Washington).

At the middle school level, we see a similar pattern. But the correlations between the derived values and the state data are lower across the board (0.18 to 0.36 in Michigan and 0.35 to

[45] Hence there are approximately 31,007 observations in Michigan corresponding the 828 districts x 19 months x 3 school levels (the number of districts varies slightly across school level given that not all districts have elementary, middle, and high schools), and 13,678 observations in Washington, corresponding to the 303 districts x 18 months x 3 school levels.

0.42 in Washington). In both states, the correlation between the R2L data and the state data is very close to zero. Finally, at the high school level, the general pattern holds. But in both states, the derived methods perform better than the R2L data when comparing them to the state data. Notably the P&L method has a far lower correlation with the state data relative to the derived methods in Michigan. But in Washington, the P&L method has a higher correlation with the state data than the other means of estimating school modality.

What explains these relatively low correlations? We cannot do much more than speculate,[46] but it is important to consider that the state surveys for much of the year ask about "planned modality" during a period when COVID surges (e.g., the Omicron wave in early 2021) could have upended modality plans. Nonetheless, none of the methods provide estimates that are particularly close to state-reported modalities. This suggests that any attempt to get at national implications of school closures—or studies that do not have official modality records—will likely suffer from substantial measurement error.[47]

5. Discussion and Conclusions

Pandemic-related remote schooling has generally been interpreted as a disaster for student achievement (Di Pietro, 2023). While that may be true on average, we know relatively little about the extent to which the effects of remote instruction varied across grade levels and schools which operated remotely or in person to different degrees during the pandemic. This is an important area for future research, as some degree of remote instruction is likely to feature in

[46] The mean characteristics across districts that agree or disagree on our various metrics are not very different from one another. These results are available upon request.

[47] In Appendix Table A1D, we report a similar correlation matrix at a national scale. Here there is no national officially collected data so we are only comparing the various derived estimates of district share remote to those collected in the R2L data. At each school level, the probability method is highly correlated with the optimal threshold method (0.87 to 0.92). The R2L data and the probability methods have the lowest correlation, 0.12 to 0.23.

the future. For example, schools are increasingly using remote instruction as a substitute for snow days or sudden school closures due to power outages. Unfortunately, research on the effects of pandemic-era remote schooling is hampered by incomplete national data on which schools were operating remotely during the 2020-21 school year.

Our analysis of different ways of using the SafeGraph mobility data suggests the use of adjusted visit counts does a good job of identifying whether schools are in session outside of a pandemic context. This is important. There are a range of possible studies—such as those examining the impact of alternative school calendars, snow days, or teacher strikes—that could benefit from an inexpensive means of identifying whether and for how long schools are operating in person or remotely or not at all.

Unfortunately, despite performing well in discriminating between schools being in and out of session in a pre-pandemic period, the use of adjusted visit counts performs relatively poorly as a means of predicting whether schools were in session *and* in person (as opposed to in session and remote) at the onset of the pandemic. The methods advanced by P&L (2021) that rely on classifying schools based on whether there is a 50% drop off in visits perform better at distinguishing between in person and remote at the onset of the pandemic. But when we focus on the 2020-21 school year in Michigan and Washington, we see a more mixed picture. The different methods produce estimates of the incidence of remote or in person schooling that are more closely aligned. Whether one or another method is more accurate depends on the weight one places on correctly identifying cases where schools are remote versus correctly identifying cases where they are in person. We also find that optimal cutoffs vary to some degree across states, suggesting a benefit to applying different methods for classifying schools in additional states where there is comprehensive state data about school modality during the 2020-21 school year. At the same time, while our two methods and the P&L (2021) method tend to be more correlated with officially reported instructional modalities from these states than the Return to Learn national tracker, all three methods perform relatively poorly with correlations ranging from 0.2 to 0.7 depending on the state, method, and grade level. Combined, these analyses suggest that there is substantial measurement error in *all* these methods that could affect the accuracy of impact estimates, though foot-traffic data appears to dominate the web-scrapping-based tracker.

Our findings illustrate the importance of carefully considering the nature of questions about remote schooling. For instance, questions about the political dynamics that led systems to open schools for in person learning require categorization of whether schools were open for the *option* of in person schooling. This requires placing a high priority on identifying that at least some students were in school, in person. Other questions might be more appropriately answered using the probabilistic approach we advance here. For instance, even when schools provided optional in person schooling, not all families would afford themselves of that option. Here it might be more important to understand the *share* of students attending schools in person when the option is available. Similarly, for questions about the efficacy (or harm) of remote schooling it is arguably better to know the proportion of students who are in a remote schooling status than whether schools were open as an in-person option. More generally, that the different measures of remote schooling during the pandemic are not in close agreement suggests a need for any research using measures of remote schooling during the pandemic to assess robustness of findings across different measures. Going forward, our findings underscore the value of data collection. Whether schools were operating remotely was a vital question during the pandemic (e.g., Goldhaber et al., 2022) and retrospectively understanding the implications of school modality on student achievement is a fundamental question for policy and research. Unfortunately, only a limited number of states collected data about school modality, making data collections like R2L and the School Data Hub important resources, and illustrating the importance of better state data collection in the future.

References

Attebury, A. (2022, June 6). DeSantis claims in-person learning saved Florida students. It's not that simple. Politico. Retrieved from https://www.politico.com/news/2022/06/08/desantis-in person-learning-florida-students-schools-00038099

Bacher-Hicks, A., Goodman, J., Green, J. G., & Holt, M. (2022). *The COVID-19 Pandemic Disrupted Both School Bullying and Cyberbullying* (No. w29590). National Bureau of Economic Research.

Baron, E. J., Goldstein, E. G., & Wallace, C. T. (2020). Suffering in silence: How COVID-19 school closures inhibit the reporting of child maltreatment. Journal of public economics, 190, 104258.

Bauer, L. (2020, May 6). The COVID-19 crisis has already left too many children hungry in America. Brookings. Retrieved June 1, 2021 from https://www.brookings.edu/blog/up-front/2020/05/06/the-covid-19-crisis-has-already-left-too-many-children-hungry-in-america/

Camp, A., Johnson, A., & Zamarro, G. (2023). Revisiting Ethnic Differences in In person Learning During the 2021-22 School Year. *Journal of School Choice*, 1-50

Camp, A. M., & Zamarro, G. (2022). Determinants of ethnic differences in school modality choices during the COVID-19 crisis. *Educational Researcher*, *51*(1), 6-16.

Campbell, A. M. (2020). An increasing risk of family violence during the Covid-19 pandemic: Strengthening community collaborations to save lives. Forensic science international: reports, 2, 100089.

Campo-Flores, A. (2021, March 17). Florida Schools Reopened Without Becoming Covid-19 Superspreaders: Schools in the state have maintained case rates lower than those in the wider community. Wall Street Journal. Retrieved from https://www.wsj.com/articles/florida-schools-reopened-without-becoming-covid-19-superspreaders-11615973402

Cantor, J., Kravitz, D., Sorbero, M., Andraka-Christou, B., Whaley, C., Bouskill, K., & Stein, B. D. (2021). Trends in visits to substance use disorder treatment facilities in 2020. *Journal of substance abuse treatment*, *127*, 108462.

Cohen, R. (August 28, 2020). Some teachers are being required to come to school - To teacher virtually. The Intercept. Retrieved from: https://theintercept.com/2020/08/28/coronavirus-schools-teachers-remote/

Cantor, J., Kravitz, D., Sorbero, M., Andraka-Christou, B., Whaley, C., Bouskill, K., & Stein, B. D. (2021). Trends in visits to substance use disorder treatment facilities in 2020. *Journal of Substance Abuse Treatment*, 108462.

Courtemanche, C. J., Le, A. H., Yelowitz, A., & Zimmer, R. (2021). *School Reopenings, Mobility, and COVID-19 Spread: Evidence from Texas* (Working Paper No. 28753). National Bureau of Economic Research. http://www.nber.org/papers/w28753

Darling-Aduana, J., Woodyard, H.T., Sass, T., Barry, S.S. (2022). Learning-Mode Choice, Student Engagement, and Achievement Growth During the COVID-19 Pandemic. CALDER Working Paper No. 260-0122.

Doyle, T., Kendrick, K., Troelstrup, T., Gumke, M., Edwards, J., Chapman, S., Propper, R., Rivkees, S., & Blackmore, C. (2021). COVID-19 in primary and secondary school settings during the first semester of school reopening—Florida, August–December 2020. *Morbidity and Mortality Weekly Report*, *70*(12), 437.

Di Pietro, G. (2023). The impact of Covid-19 on student achievement: Evidence from a recent meta-analysis. *Educational Research Review*, 100530.

Eggert, D. (2021, February 23). More than 80% of Michigan schools offering some in person learning. Detroit Free Press. Retrieved from: https://www.freep.com/story/news/education/2021/02/23/more-than-80-percent-schools-return-in person-learning/4555057001/

Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual review of sociology*, *40*, 31-53.

Furfaro, H., Villa, M., & Bazzaz, D. (2021, April 19). Politics, race were key factors when Washington schools reopened for in person learning during pandemic. The Seattle Times. Retrieved from: https://www.seattletimes.com/education-lab/politics-race-determined-which-students-made-it-back-to-school-buildings-during-the-pandemic/

Golberstein, E., Wen, H., & Miller, B. F. (2020). Coronavirus disease 2019 (COVID-19) and mental health for children and adolescents. JAMA pediatrics, 174(9), 819-820.

Goldhaber, D., Imberman, S. A., Strunk, K. O., Hopkins, B. G., Brown, N., Harbatkin, E., & Kilbride, T. (2022). To what extent does in-person schooling contribute to the spread of Covid-19? Evidence from Michigan and Washington. *Journal of Policy Analysis and Management*, *41*(1), 318-349.

Goldhaber, D., Kane, T. J., McEachin, A., & Morton, E. (2022). A Comprehensive Picture of Achievement across the COVID-19 Pandemic Years: Examining Variation in Test Levels and Growth across Districts, Schools, Grades, and Students. Working Paper No. 265-0522. *National Center for Analysis of Longitudinal Data in Education Research (CALDER)*.

Goldhaber, D., Kane, T. J., McEachin, A., Morton, E., Patterson, T., & Staiger, D. O. (2023). The educational consequences of remote and hybrid instruction during the pandemic. *American Economic Review: Insights*, *5*(3), 377-392.

Grossmann, M., Reckhow, S., Strunk, K. O., & Turner, M. (2021). All States Close but Red Districts Reopen: The Politics of In person Schooling During the COVID-19 Pandemic. *Educational Researcher*, *50*(9), 637-648. https://doi.org/10.3102/0013189X211048840

Graves, J. (2010). The academic impact of multi-track year-round school calendars: A response to school overcrowding. *Journal of urban Economics*, *67*(3), 378-391.

Goodman, J. (2014). Flaking out: Student absences and snow days as disruptions of instructional time (No. w20221). National Bureau of Economic Research.

Halloran, C., Jack, R., Okun, J. C., & Oster, E. (2021). Pandemic schooling mode and student test scores: Evidence from us states (No. w29497). National Bureau of Economic Research.

Hansen, B., Sabia, J. J., & Schaller, J. (2022). *In person Schooling and Youth Suicide: Evidence from School Calendars and Pandemic School Closures* (No. w30795). National Bureau of Economic Research.

Hansen, B., Sabia, J. J., & Schaller, J. (2021). Zooming in on Schools and the Gendered Impact of the Covid-19 Pandemic [Manuscript submitted for publication].

Hawrilenko, M., Kroshus, E., Tandon, P., & Christakis, D. (2021). The association between school closures and child mental health during COVID-19. *JAMA network open*, *4*(9), e2124092-e2124092.

Hemphill, A. A., & Marianno, B. D. (2021). Teachers' unions, collective bargaining, and the response to COVID-19. *Education Finance and Policy*, *16*(1), 170-182.

Hopkins, B., Kilbride, T., & Strunk, K.O. (May, 2021). Instructional Delivery Under Michigan Districts' Extended COVID-19 Learning Plans – May Update. The Education Policy Innovation Collaborative, Michigan State University. East Lansing, MI. https://epicedpolicy.org/wp-content/uploads/2021/05/EPIC_ECOL_report_May2021.pdf

Hosmer DW, Lemeshow S (2000). Applied logistic regression, 2nd ed. Wiley, pp 156–164

Huang, X., Lu, J., Gao, S., Wang, S., Liu, Z., & Wei, H. (2022). Staying at home is a privilege: evidence from fine-grained mobile phone location data in the United States during the COVID-19 pandemic. *Annals of the American Association of Geographers*, *112*(1), 286-305.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Hopkins, B., Strunk, K.O., & Kilbride, T. (2022) "Differential Student Uptake of In person Instruction During the 2020-21 School Year: Evidence from Michigan." https://epicedpolicy.org/wp-content/uploads/2023/03/COVIDUptake_WorkingPaper_Jan2023.pdf

Huntington-Klein, N. (2020). *Walking in (The University Of) Memphis: Which College Campuses Opened in Fall 2020?*. Seattle University working paper.

Jack, R., & Oster, E. (2023). COVID-19, School Closures, and Outcomes. *Journal of Economic Perspectives*, *37*(4), 51-70.

Kam, D. (November 18, 2020). Florida education chief says distance learning to continue in 2nd half of school year because of pandemic. The Daytona Beach News-Journal. Retrieved from: https://www.news-journalonline.com/story/news/state/2020/11/18/florida-education-chief-says-pandemic-means-state-continue-distance-learning-coronavirus-covid/3768853001/

Kilbride, T., Hopkins, B., Strunk, K.O., & Imberman, S. (December, 2021). K-8 Student Achievement and Achievement Gaps on Michigan's 2020-21 Benchmark and Summative Assessments. The Education Policy Innovation Collaborative, Michigan State University. East Lansing, MI. https://epicedpolicy.org/k-8-student-achievement-and-achievement-gaps-on-michigans-2020-21-benchmark-and-summative-assessments/

Kogan, V., & Lavertu, S. (2021). How the covid-19 pandemic affected student learning in ohio: Analysis of spring 2021 Ohio state tests. *Ohio State University, John Glenn College of Public Affairs*, *28*, 2021-10.

Kurmann, A. & Lale, E. (January, 2022). School Closures, Effective In person Learning during COVID-19: When, Where, and for Whom. IZA Institute of Labor Economics Discussion Paper No. 14984. https://docs.iza.org/dp14984.pdf

Lee, S. J., Ward, K. P., Chang, O. D., & Downing, K. M. (2021). Parenting activities and the transition to home-based education during the COVID-19 pandemic. Children and Youth Services Review, 122, 105585.

Marcotte, D. E., & Hemelt, S. W. (2008). Unscheduled school closings and student performance. *Education Finance and Policy*, *3*(3), 316-338.

McEachin, A., & Atteberry, A. (2017). The impact of summer learning loss on measures of school performance. *Education Finance and Policy*, *12*(4), 468-491.

McMullen, S. C., & Rouse, K. E. (2012). The impact of year-round schooling on academic achievement: Evidence from mandatory school calendar conversions. *American Economic Journal: Economic Policy*, *4*(4), 230-52.

Morton, E. (2021). Effects of four-day school weeks on school finance and achievement: Evidence from Oklahoma. *Educational Researcher*, *50*(1), 30-40.

Parks, S. E., Zviedrite, N., Budzyn, S. E., Panaggio, M. J., Raible, E., Papazian, M., ... & Barrios, L. C. (2021). COVID-19–Related School Closures and Learning Modality Changes—United States, August 1–September 17, 2021. *Morbidity and Mortality Weekly Report*, *70*(39), 1374.

Parolin, Z., & Lee, E. K. (2021). Large socio-economic, geographic and demographic disparities exist in exposure to school closures. *Nature human behaviour*, *5*(4), 522-528.

Peele, H., & Riser-Kositsky, M. (2020). Map: Coronavirus and school closures in 2019-2020. *Education Week*.

Reardon, S., Kalogrides, D., Ho, A., Shear, B., Fahle, E., Jang, H., & Chavez, B. (2021). Stanford Education Data Archive (SEDA).

Sachs, E., Miller, L., & Schueler, B. (2022). Virginia School Division Operations During SY 2020-21: In person Learning. University of Virginia, Ed Policy Works. Retrieved from: https://education.virginia.edu/sites/default/files/uploads/epw/Division%20Operations%20In person%20Learning%2020220925_0.pdf

SafeGraph. (n.d.-a). Known Issues or Data Artifacts. Retrieved from: https://docs.safegraph.com/docs/known-issues-or-data-artifacts

SafeGraph. (n.d.-b).Privacy Policy. Retrieved from: https://www.safegraph.com/privacy-policy

Solochek, J.. (2020, November 30). Florida makes it official: Remote learning will continue next semester. Tampa Bay Times. Retrieved from https://www.tampabay.com/news/education/2020/11/30/florida-makes-it-official-remote-learning-will-continue-next-semester/

Squire, R. (2019). What About Bias in the SafeGraph Dataset?: Quantifying Sampling Bias in SafeGraph Patterns. Retrieved from https://www.safegraph.com/blog/what-about-bias-in-the-safegraph-dataset

Thompson, P. N. (2021). Is four less than five? Effects of four-day school weeks on student achievement in Oregon. *Journal of Public Economics*, *193*, 104308.

Ujifusa, A. (2020). States Ordering Schools to Close in Response to Coronavirus. EdWeek. Retrieved from: https://www.edweek.org/education/states-ordering-schools-to-close-in-response-to-coronavirus/2020/03

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2020). Education: From disruption to recovery.

U.S. Department of Education, National Center for Education Statitics. (2021). *Digest of Education Statistics*, 2019 (NCES 2021-009), Table 105.50.

West, M. R. & Lake, R. (2021, August). How has the pandemic affected students' social-emotional well-being? A review of the evidence to date. Center on Reinventing Public Education. https://www.crpe.org/publications/how-has-pandemic-affected-students-social-emotional-well-being-review-evidence-date.

Woody, S., Tec, M., Dahan, M., Gaither, K., Lachmann, M., Fox, S., Meyers, L., & Scott, J. (2020). Projections for first-wave COVID-19 deaths across the US using social-distancing measures derived from mobile phones. *Medrxiv*.

Yan, Y., Bayham, J., Richter, A., & Fenichel, E. P. (2021). Risk compensation and face mask mandates during the COVID-19 pandemic. *Scientific reports*, 11(1), 1-11.

Yan, Y., Malik, A. A., Bayham, J., Fenichel, E. P., Couzens, C., & Omer, S. B. (2021). Measuring voluntary and policy-induced social distancing behavior during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences*, *118*(16).

Tables and Figures

Table 1. Summary statistics of schools in analytic sample

Panel A	Mean School Characteristics		
	SafeGraph sample	Unmatched sample	
Total enrollment	570	408	
% Under-represented minority	44.9	50.1	
% FRL	49.5	48.5	
Urban	26.9%	29.1%	
Suburban	32.8%	25.7%	
Town	12.6%	13.1%	
Rural	27.7%	32.1%	
Unique Schools	78,905	11,788	
Panel B	**SafeGraph Visits**		
	1st Week October 2019	1st Week October 2020	% Change
Q1 Total enrollment (lowest)	14.9	10.1	-32.2%
Q4 Total enrollment	98.7	44.0	-55.4%
Q1 % FRL (lowest)	51.9	22.1	-57.4%
Q4 %FRL	21.1	10.3	-51.2%
Q1 % URM (lowest)	44.9	28.8	-35.9%
Q4 % URM	23.0	7.8	-66.1%
Urban	33.6	11.7	-65.2%
Suburban	52.4	22.6	-56.9%
Town	41.6	26.8	-35.6%
Rural	37.4	24.2	-35.3%
Elementary	16.5	9.7	-41.2%
Middle	54.0	25.5	-52.8%
High School	99.5	46.4	-53.4%
Other	22.4	13.3	-40.6%

Note: Non-matches have a high rate of Alternative School, Career and Technical School, Special Education School types. The matched sample is comprised of 93.1% traditional schools whereas only 37.8% of unmatched are traditional types. Non-traditional schools include youth juvenile detention education programs, special education only programs, etc.

Table 2. District modality snapshot (1st week Oct. 2020 instructional setting)

	Michigan				Washington			
	All	In Person	Hybrid	Remote	All	In Person	Hybrid	Remote
School Characteristics								
Enrollment	460.5	433.6	522.5	479.8	519.0	299.9	340.8	551.8
% FRL	51.5	51.3	43.3	63.3	45.3	42.8	47.8	45.0
Urban	17.7	10.7	16.1	47.7	30.6	0.0	2.8	35.3
Suburban	39.1	31.7	53.2	47.0	35.0	24.0	14.2	38.9
Town	14.4	19.0	10.4	2.5	13.9	18.0	35.1	10.4
Rural	28.8	38.7	20.3	2.8	20.5	58.0	47.9	15.5
Unique Schools	2,496	1,551	547	398	2,034	80	257	1,697

Table 3. Alignment between R2L and Data Hub modality data

	Panel A: National		
	DataHub In Person	DataHub Hybrid	DataHub Remote
R2L In Person	85.24%	11.41%	3.36%
R2L Hybrid	27.79%	59.44%	12.77%
R2L Remote	4.59%	12.97%	82.44%
	Panel B: Michigan		
	DataHub In Person	DataHub Hybrid	DataHub Remote
R2L In Person	76.00%	21.15%	2.85%
R2L Hybrid	24.10%	69.59%	6.31%
R2L Remote	8.32%	23.13%	68.56%
	Panel C: Washington		
	DataHub In Person	DataHub Hybrid	DataHub Remote
R2L In Person	91.37%	8.50%	0.13%
R2L Hybrid	2.52%	70.85%	26.62%
R2L Remote	0.04%	1.97%	98.00%

Note: The Data Hub data are recorded at both the school and district levels, and across different time intervals (monthly, bi-weekly, weekly, and daily). We assigned all schools within a district to the same modality and we converted the time interval to a weekly measure for comparative purposes. The national modality trackers overlap for 43,886 unique schools. The denominator is based on the row.

Table 4. Pre-pandemic estimates of schools being in session (Nationwide)

		Elementary	**Middle**	**High School**
		(1)	(2)	(3)
Adjusted Count		1.083***	0.862***	0.874***
		(0.002)	(0.002)	(0.002)
Enrollment Q2		-1.675***	-1.730***	-1.310***
		(0.006)	(0.012)	(0.011)
Enrollment Q3		-1.791***	-2.570***	-2.434***
		(0.006)	(0.012)	(0.013)
Enrollment Q4		-1.757***	-2.930***	-4.257***
		(0.008)	(0.010)	(0.010)
URM Q2		-0.365***	0.108***	0.121***
		(0.006)	(0.011)	(0.010)
URM Q3		-0.320***	0.229***	0.375***
		(0.006)	(0.012)	(0.010)
URM Q4		-0.142***	0.732***	1.035***
		(0.008)	(0.014)	(0.012)
FRPL Q2		-0.432***	-0.226***	-0.303***
		(0.006)	(0.011)	(0.010)
FRPL Q3		-0.283***	-0.195***	-0.045***
		(0.006)	(0.011)	(0.010)
FRPL Q4		-0.137***	-0.348***	0.118***
		(0.007)	(0.014)	(0.012)
Urban		0.091***	0.105***	0.398***
		(0.002)	(0.004)	(0.004)
Rural		0.652***	0.407***	0.362***
		(0.003)	(0.004)	(0.004)
Town		0.081***	0.037***	0.080***
		(0.003)	(0.005)	(0.004)
Adjusted count	Enrollment Q2	0.348***	0.303***	0.122***
		(0.002)	(0.003)	(0.002)
	Enrollment Q3	0.300***	0.415***	0.265***
		(0.002)	(0.003)	(0.003)
	Enrollment Q4	0.200***	0.369***	0.417***
		(0.002)	(0.002)	(0.002)
	URM Q2	0.131***	-0.002	-0.017***
		(0.002)	(0.002)	(0.002)
	URM Q3	0.155***	0.030***	-0.019***
		(0.002)	(0.003)	(0.002)
	URM Q4	0.115***	-0.015***	-0.071***
		(0.002)	(0.003)	(0.003)

	FRPL Q2	0.099***	0.057***	0.066***
		(0.002)	(0.002)	(0.002)
	FRPL Q3	0.074***	0.086***	0.061***
		(0.002)	(0.003)	(0.002)
	FRPL Q4	0.048***	0.169***	0.091***
		(0.002)	(0.003)	(0.003)
Num.Obs.		4,659,094	1,509,746	1,625,231
R2		0.509	0.505	0.43
R2 Adj.		0.509	0.505	0.43
AIC		9,918,773	3,240,892	4,011,111
BIC		9,919,094	3,241,186	4,011,406
RMSE		0.33	0.33	0.36
Std.Errors		IID	IID	IID

Notes: The dependent variable is a dichotomous indicator for the school-week being in session (versus out-of-session). Not reported: coefficient on the seasonality control: month-of-year average visits in 2018.

Table 5. AUC comparison (Michigan & Washington)

Michigan			
	Elementary	Middle	High
Our Method	0.741	0.809	0.775
P&L	0.745	0.815	0.786
Washington			
	Elementary	Middle	High
Our Method	0.723	0.807	0.781
P&L	0.745	0.838	0.799

This table shows Area Under Curve values for ROC curves, comparing actual in person schooling against predicted classifications of in person schooling over the full range of potential cutoffs. Our method is an out-of-sample prediction, using a model fitted using pre-pandemic data to classify in person schooling during the pandemic. P&L classification is done across the full possible range of thresholds, not just 50% as in the original paper.

Table 6. Comparison of different ways of predicting school modality in 2020-21 [Our Method / P&L]

	Remote Weight	*Our Method* Threshold	Actually Remote	Actually In Person	*P&L (With Varying Threshold)* Threshold	Actually Remote	Actually In Person
		Panel A Michigan					
Elementary	75%	0.96	93.64%	38.15%	4.79	99.39%	10.75%
	67%	0.94	88.28%	52.07%	0.70	81.09%	59.49%
	50%	0.89	77.20%	64.99%	0.55	72.86%	71.97%
	33%	0.78	55.23%	78.64%	0.39	56.74%	83.31%
	25%	0.01	0.93%	99.79%	0.31	46.98%	87.52%
Middle	75%	0.88	94.77%	59.59%	0.52	89.31%	65.00%
	67%	0.88	94.77%	59.59%	0.40	85.32%	74.28%
	50%	0.83	88.96%	68.67%	0.39	84.76%	75.09%
	33%	0.76	79.07%	75.33%	0.21	69.74%	84.34%
	25%	0.66	64.85%	81.51%	0.21	69.74%	84.34%
High	75%	0.88	95.91%	47.23%	0.56	90.91%	57.49%
	67%	0.81	90.80%	59.85%	0.42	87.50%	66.66%
	50%	0.80	89.55%	61.47%	0.41	87.27%	67.00%
	33%	0.64	71.59%	73.40%	0.22	73.18%	76.23%
	25%	0.03	4.89%	98.63%	0.11	53.86%	84.40%
		Panel B Washington					
Elementary	75%	0.96	92.27%	42.63%	1.02	89.54%	47.11%
	67%	0.95	89.47%	49.74%	0.93	87.79%	51.84%
	50%	0.92	79.20%	62.11%	0.62	74.96%	69.21%
	33%	0.69	38.64%	85.00%	0.47	61.03%	80.26%
	25%	0.05	3.04%	100.00%	0.35	45.46%	86.84%
Middle	75%	0.87	96.03%	56.84%	0.46	89.18%	64.21%
	67%	0.85	94.68%	60.00%	0.29	80.90%	82.11%
	50%	0.74	80.98%	76.84%	0.23	75.91%	88.42%
	33%	0.69	74.81%	81.05%	0.23	75.91%	88.42%
	25%	0.69	74.81%	81.05%	0.16	64.58%	93.68%
High	75%	0.81	94.29%	56.57%	0.28	85.39%	72.00%
	67%	0.80	94.09%	57.14%	0.27	84.38%	74.29%
	50%	0.75	90.49%	62.86%	0.25	83.38%	75.43%
	33%	0.61	74.98%	72.57%	0.15	71.47%	84.00%
	25%	0.23	25.43%	93.71%	0.10	55.66%	89.71%

Threshold for Our Method is on a probability scale, where 1 would be a predicted 100% probability of being in person. Scale for P&L is on a "current traffic as a share of pre-pandemic traffic" scale, where 1 would be no change.

Table 7. Comparison of estimates of remote status (Michigan & Washington)

		Panel A: Michigan				
Elementary Level		Reported to MI SEA[a]	Probability Method[b]	Optimal Threshold[c]	P&L Method[d]	R2L[e]
	Reported to MI SEA	1.00				
	Probability Method	0.48	1.00			
	Optimal Threshold	0.54	0.90	1.00		
	P&L Method	0.53	0.50	0.59	1.00	
	R2L	0.11	0.43	0.46	0.51	1.00
Middle School Level		Reported to MI SEA[a]	Probability Method[b]	Optimal Threshold[c]	P&L Method[d]	R2L[e]
	Reported to MI SEA	1.00				
	Probability Method	0.18	1.00			
	Optimal Threshold	0.28	0.79	1.00		
	P&L Method	0.36	0.35	0.47	1.00	
	R2L	0.02	0.20	0.27	0.42	1.00
High School Level		Reported to MI SEA[a]	Probability Method[b]	Optimal Threshold[c]	P&L Method[d]	R2L[e]
	Reported to MI SEA	1.00				
	Probability Method	0.65	1.00			
	Optimal Threshold	0.74	0.83	1.00		
	P&L Method	0.36	0.45	0.52	1.00	
	R2L	0.06	0.14	0.18	0.47	1.00

a. The Michigan State Education Agency provided modality data reported by the schools – schools provided monthly reports on their remote or in person status.

b. For each school, we compute the probability a school is remote based on the logit model in Equation (1) and the coefficient estimates from Table 4 applied to the characteristics of each school in each district.

c. Optimal thresholds are from Table 6. We use the threshold from the associated with our method at various school levels, and with remote and in person schools weighted equally.

d. The Parolin and Lee (P&L) Method uses a year-over-year measure to determine the remote status of school. Their method is to label a school as remote in a given month if it exhibits a 50% drop-off in visits from Month X in Year Y to Month X in year Y+1.

e. R2L assumes a district operates under one modality, and scrapes data from district websites to classify whether a school is in person or remote.

		Panel B: Washington				
Elementary Level		Reported to WA SEA[a]	Probability Method[b]	Optimal Threshold[c]	P&L Method[d]	R2L[e]
	Reported to WA SEA	1.00				
	Probability Method	0.59	1.00			
	Optimal Threshold	0.70	0.91	1.00		
	P&L Method	0.56	0.51	0.65	1.00	
	R2L	0.18	0.52	0.55	0.66	1.00
Middle School Level		Reported to WA SEA[a]	Probability Method[b]	Optimal Threshold[c]	P&L Method[d]	R2L[e]
	Reported to WA SEA	1.00				
	Probability Method	0.35	1.00			
	Optimal Threshold	0.42	0.90	1.00		
	P&L Method	0.38	0.21	0.29	1.00	
	R2L	0.07	0.14	0.16	0.66	1.00
High School Level		Reported to WA SEA[a]	Probability Method[b]	Optimal Threshold[c]	P&L Method[d]	R2L[e]
	Reported to WA SEA	1.00				
	Probability Method	0.19	1.00			
	Optimal Threshold	0.23	0.89	1.00		
	P&L Method	0.36	0.19	0.32	1.00	
	R2L	0.11	0.15	0.25	0.60	1.00

a. The Washington State Education Agency provided modality data reported by the schools – schools provided monthly reports on their remote or in person status.

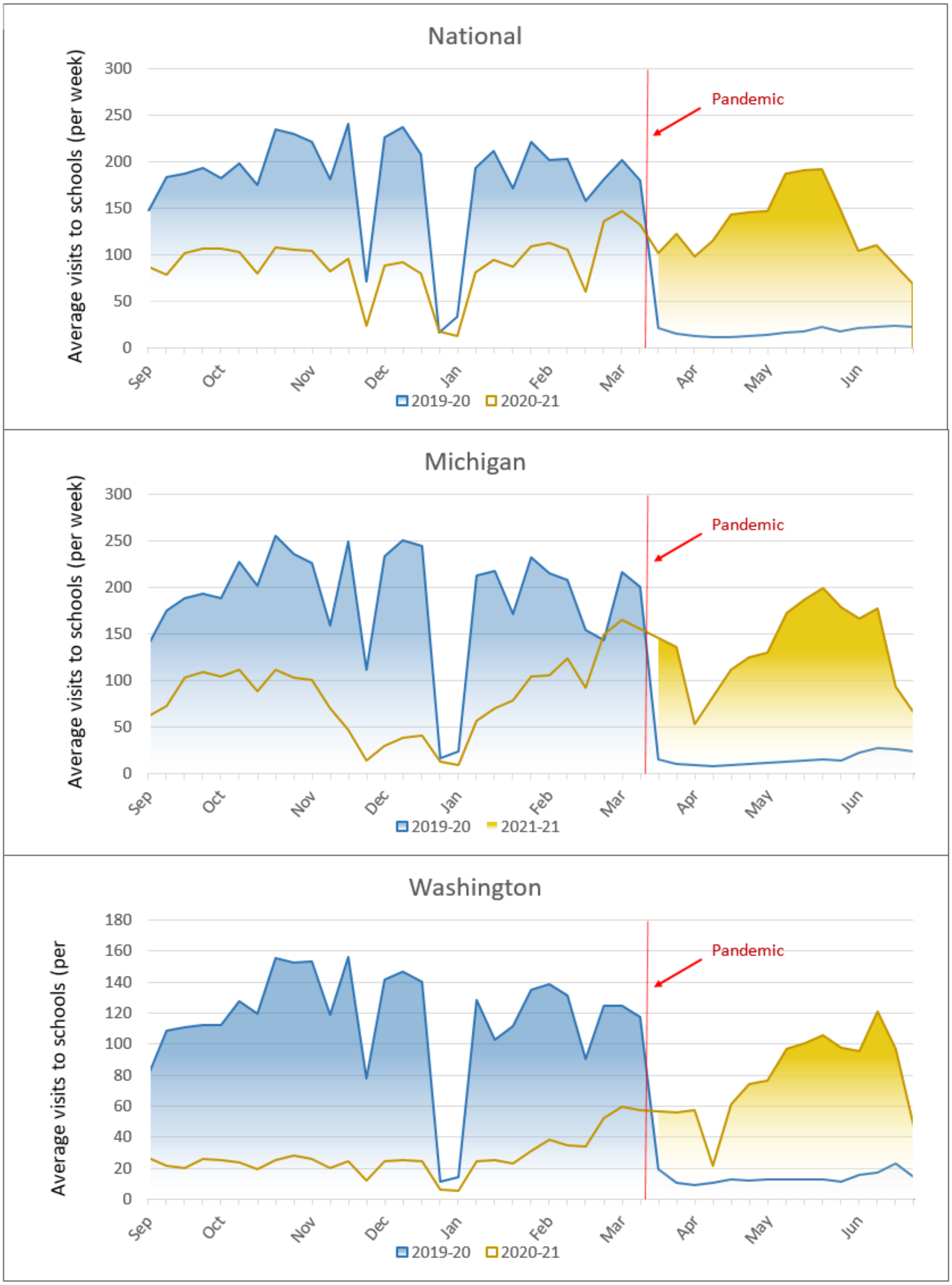
b. For each school, we compute the probability a school is remote based on the logit model in Equation (1) and the coefficient estimates from Table 4 applied to the characteristics of each school in each district.

c. Optimal thresholds are from Table 6. We use the threshold from the associated with our method at various school levels, and with remote and in person schools weighted equally.

d. The Parolin and Lee (P&L) Method uses a year-over-year measure to determine the remote status of school. Their method is to label a school as remote in a given month if it exhibits a 50% drop-off in visits from Month X in Year Y to Month X in year Y+1.

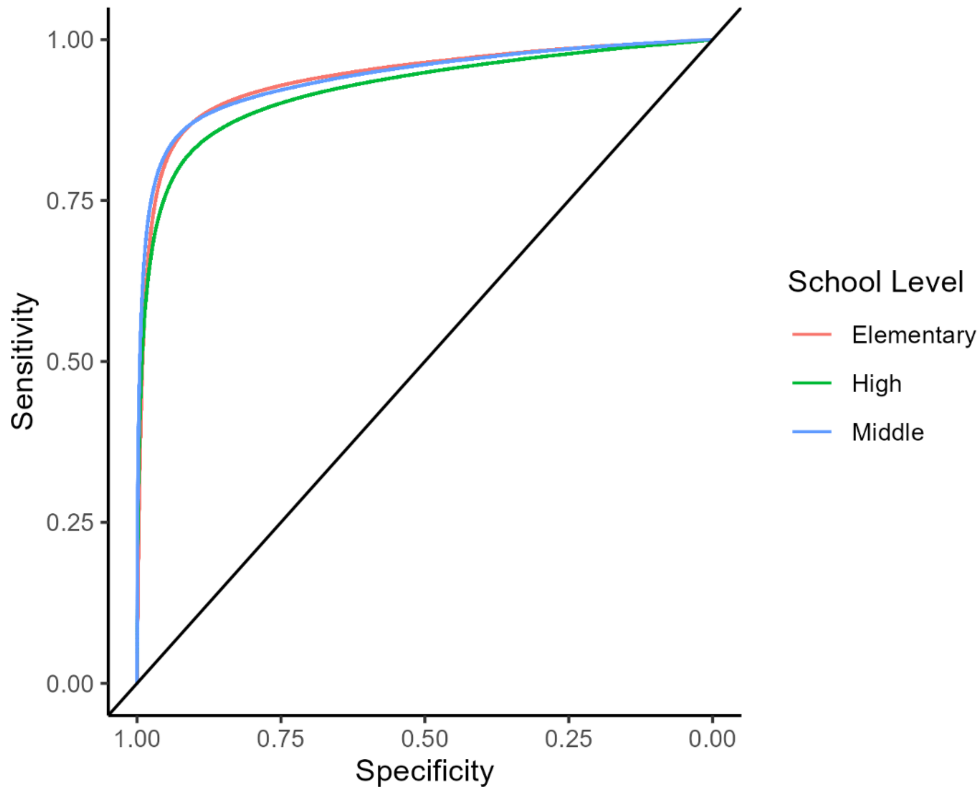
e. R2L assumes a district operates under one modality, and scrapes data from district websites to classify whether a school is in person or remote.

Figure 1. Comparing school visits per week before and during the pandemic



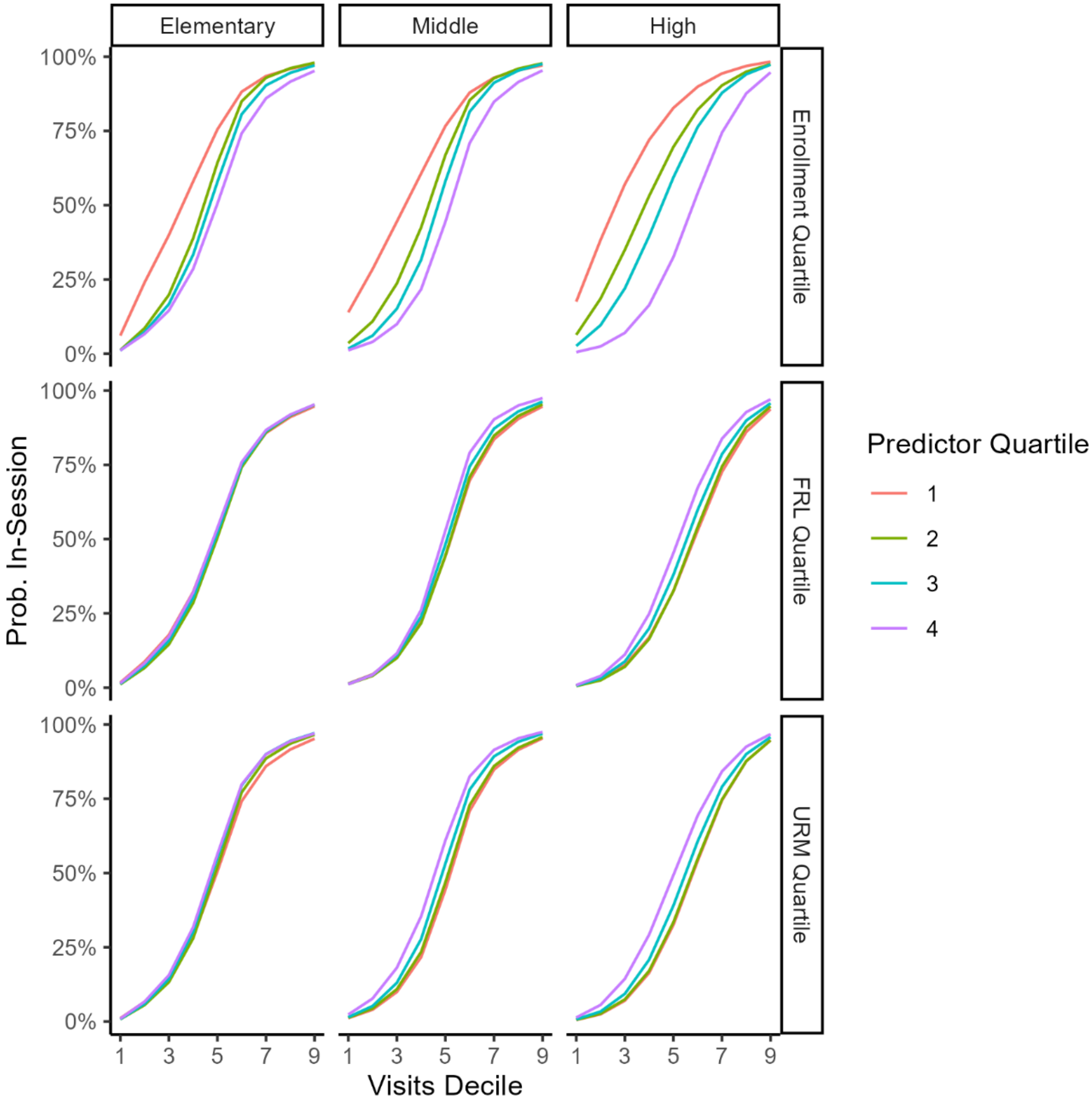
Because days shift slightly between years we begin with the first week in September that has a majority of days in that particular month (i.e., the week beginning September 2, 2019 and August 31, 2020. This was also done for assignment of weeks to months. Figure based on matched sample of 84,101 unique schools.

Figure 2. ROC curves showing national in-sample prediction in pre-pandemic period



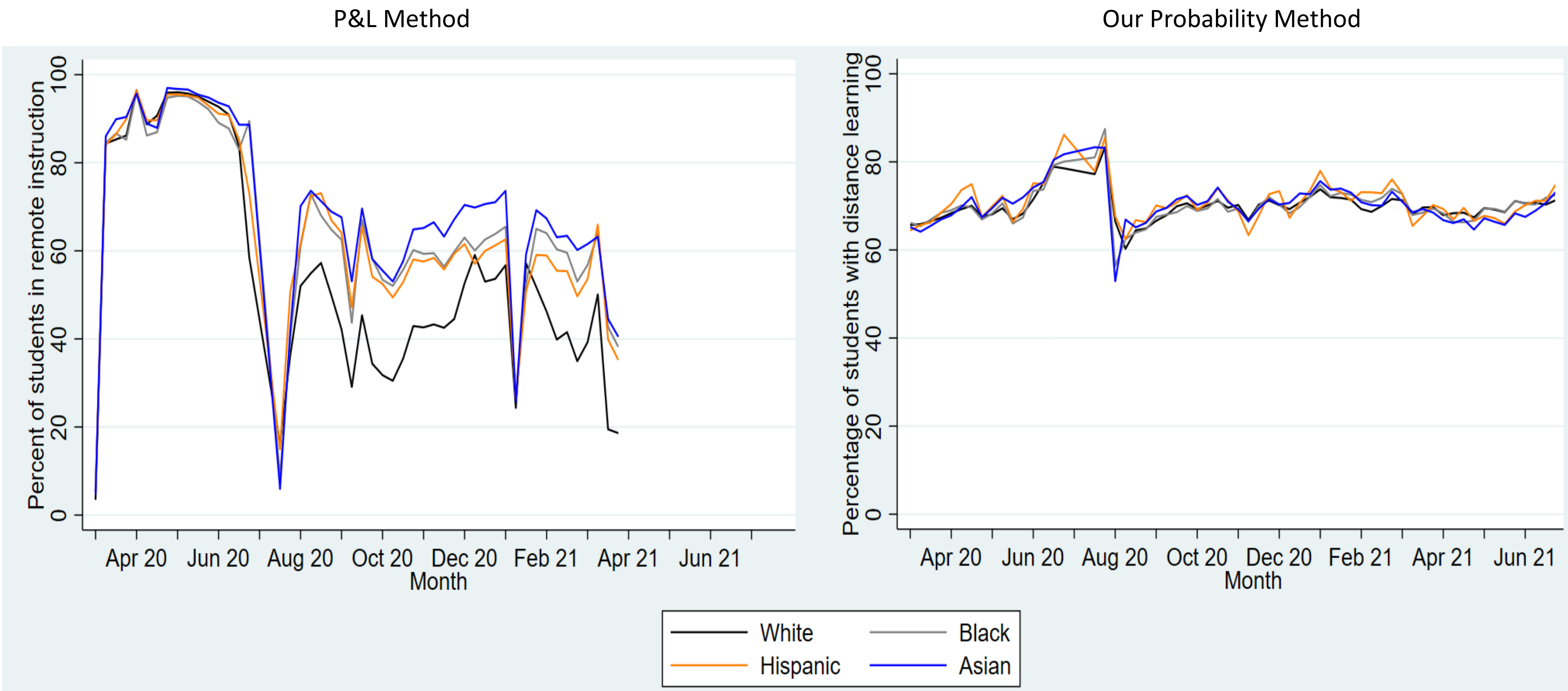
Each curve compares the predicted in-session vs. out-of-session classification of each observation in the pre-pandemic period using national data against actual in-session status, across a range of classification cutoffs. AUC values for Elementary, Middle, and High school .940, .941, and .922, respectively.

Figure 3. National comparison of predicted probability of in session



This figure shows the predicted probability of being in session based on the all-states regression model. The x-axis is the decile of adjusted visits within school level, calculated across all schools of that type, all pre-pandemic weeks, and both in-session and out-of-session days. Each line shows how the predicted probability of being in session increases for higher visit-count deciles. Each row of panels performs this calculation while varying one covariate. In the top row, for example, we see that schools in lower enrollment quartiles are more likely to be in session at every quartile of visit counts, and also that the probability increases with visit counts more quickly.

Figure 4. P&L vs. Probability method for determining proportion of students in distance learning (by ethnicity)



These figures show the proportion of students in each demographic predicted to be in remote schooling. The first panel uses the P&L 50% dropoff rule to classify students. The second panel uses predicted probabilities of being remote from Equation (1), averaging them together, weighted by enrollment by demographic. Figure based on matched sample of 84,101 unique schools.

Appendix

Table A1A. Marginal effects (Elementary School)

	Elementary					
	Michigan			Washington		
	(1)	(2)	(3)	(10)	(11)	(12)
asinh(adjusted count)	1.427***	1.549***	1.285***	1.441***	1.711***	0.994***
	-0.003	-0.004	-0.01	-0.004	-0.005	-0.014
urban		0.031**	0.027**		0.005	−0.004
		-0.012	-0.013		-0.012	-0.013
rural		0.952***	0.863***		1.524***	1.387***
		-0.014	-0.014		-0.018	-0.018
town		0.279***	0.308***		0.171***	0.150***
		-0.015	-0.015		-0.018	-0.018
sy2018_month_average		−0.013***	−0.013***		−0.014***	−0.014***
		0	0		0	0
ccd_tot_enrollment_q2		−0.525***	−1.089***		−0.726***	−1.589***
		-0.011	-0.027		-0.017	-0.035
ccd_tot_enrollment_q3		−0.858***	−2.247***		−1.015***	−2.144***
		-0.013	-0.036		-0.017	-0.035
ccd_tot_enrollment_q4		−0.820***	−0.668***		−1.167***	−2.020***
		-0.02	-0.049		-0.02	-0.048
URM_quartile2		0.017	−0.584***		−0.836***	−1.221***
		-0.012	-0.029		-0.025	-0.04
URM_quartile3		0.182***	−0.561***		−0.624***	−0.920***
		-0.015	-0.035		-0.026	-0.042
URM_quartile4		0.162***	−1.415***		−0.278***	−0.208***
		-0.017	-0.041		-0.028	-0.046
frl_quartile2		−0.214***	−0.229***		−0.270***	−0.473***
		-0.015	-0.038		-0.015	-0.033
frl_quartile3		−0.191***	0.189***		−0.407***	−1.062***
		-0.015	-0.034		-0.016	-0.036
frl_quartile4		−0.117***	0.212***		−0.496***	−1.376***
		-0.016	-0.038		-0.019	-0.042
asinh(count_adjust) × ccd_tot_enrollment_q2			0.196***			0.380***
			-0.008			-0.012
asinh(count_adjust) × ccd_tot_enrollment_q3			0.430***			0.460***
			-0.01			-0.012
asinh(count_adjust) × ccd_tot_enrollment_q4			−0.015			0.373***
			-0.014			-0.015
asinh(count_adjust) × URM_quartile2			0.195***			0.339***
			-0.009			-0.016
asinh(count_adjust) × URM_quartile3			0.243***			0.283***
			-0.01			-0.016
asinh(count_adjust) × URM_quartile4			0.509***			0.132***
			-0.012			-0.017
asinh(count_adjust) × frl_quartile2			0			0.092***
			-0.011			-0.011
asinh(count_adjust) × frl_quartile3			−0.131***			0.265***
			-0.01			-0.012
asinh(count_adjust) × frl_quartile4			−0.112***			0.347***
			-0.011			-0.014

Num.Obs.	177373	177373	177373	140282	140282	140282
R2	0.511	0.541	0.547	0.503	0.559	0.565
R2 Adj.	0.511	0.541	0.547	0.503	0.559	0.565
AIC	375828.7	353477.8	348446.9	302500.9	268398.3	264899.1
BIC	375848.9	353629.1	348689	302520.6	268546.1	265135.5
RMSE	0.32	0.31	0.31	0.33	0.31	0.3
Std.Errors	IID	IID	IID	IID	IID	IID

Table A1B. Marginal effects (Middle School)

	Middle					
	Michigan			Washington		
	(1)	(2)	(3)	(10)	(11)	(12)
asinh(count_adjust)	1.056***	1.182***	0.778***	1.063***	1.280***	0.834***
	-0.004	-0.005	-0.014	-0.005	-0.006	-0.024
urban		0.213***	0.237***		0.234***	0.211***
		-0.023	-0.024		-0.023	-0.024
rural		0.080***	0.111***		0.964***	0.919***
		-0.022	-0.022		-0.034	-0.034
town		0.195***	0.202***		0.161***	0.193***
		-0.024	-0.024		-0.029	-0.03
sy2018_month_average		−0.010***	−0.010***		−0.011***	−0.012***
		0	0		0	0
ccd_tot_enrollment_q2		−0.720***	−1.045***		−1.235***	−4.386***
		-0.024	-0.053		-0.041	-0.151
ccd_tot_enrollment_q3		−1.197***	−3.234***		−1.028***	−1.118***
		-0.024	-0.064		-0.035	-0.064
ccd_tot_enrollment_q4		−1.500***	−2.706***		−1.571***	−1.902***
		-0.025	-0.055		-0.035	-0.062
URM_quartile2		−0.064***	−0.305***		−1.161***	−0.803***
		-0.019	-0.049		-0.045	-0.073
URM_quartile3		−0.043*	−0.086		−1.043***	−1.601***
		-0.024	-0.066		-0.045	-0.076
URM_quartile4		0.168***	−0.108		−0.887***	−1.553***
		-0.029	-0.074		-0.049	-0.09
frl_quartile2		0.151***	−0.150**		−0.121***	−0.887***
		-0.022	-0.058		-0.028	-0.065
frl_quartile3		0.323***	−0.053		−0.304***	−1.447***
		-0.023	-0.058		-0.03	-0.066
frl_quartile4		0.539***	−0.637***		0.190***	−0.243***
		-0.027	-0.075		-0.038	-0.083
asinh(count_adjust) × ccd_tot_enrollment_q2			0.137***			0.806***
			-0.014			-0.038
asinh(count_adjust) × ccd_tot_enrollment_q3			0.554***			0.043**
			-0.016			-0.018
asinh(count_adjust) × ccd_tot_enrollment_q4			0.356***			0.114***
			-0.013			-0.017
asinh(count_adjust) × URM_quartile2			0.058***			0.037
			-0.011			-0.024
asinh(count_adjust) × URM_quartile3			0.019			0.263***
			-0.015			-0.024
asinh(count_adjust) × URM_quartile4			0.085***			0.290***
			-0.018			-0.028
asinh(count_adjust) × frl_quartile2			0.070***			0.197***
			-0.013			-0.015
asinh(count_adjust) × frl_quartile3			0.082***			0.291***
			-0.013			-0.017
asinh(count_adjust) × frl_quartile4			0.307***			0.118***
			-0.018			-0.022

Num.Obs.	59791	59791	59791	42832	42832	42832
R2	0.449	0.489	0.497	0.457	0.521	0.531
R2 Adj.	0.449	0.489	0.497	0.457	0.521	0.531
AIC	142829.2	132517.1	130465.7	100804.2	89048.3	87079.5
BIC	142847.2	132652.1	130681.6	100821.6	89178.3	87287.5
RMSE	1.056***	1.182***	0.778***	1.063***	1.280***	0.834***
Std.Errors	IID	IID	IID	IID	IID	IID

Table A1C. Marginal effects (High School)

	High					
	Michigan			Washington		
	(1)	(2)	(3)	(10)	(11)	(12)
asinh(count_adjust)	0.781***	1.085***	0.812***	0.656***	1.046***	0.781***
	-0.003	-0.004	-0.012	-0.004	-0.005	-0.016
urban		0.467***	0.569***		−0.076***	−0.097***
		-0.022	-0.023		-0.022	-0.022
rural		0.442***	0.467***		0.206***	0.272***
		-0.019	-0.019		-0.027	-0.027
town		0.287***	0.346***		−0.075***	−0.053*
		-0.021	-0.021		-0.027	-0.028
sy2018_month_average		−0.008***	−0.008***		−0.007***	−0.007***
		0	0		0	0
ccd_tot_enrollment_q2		−0.661***	−0.944***		−1.252***	−0.800***
		-0.021	-0.051		-0.034	-0.069
ccd_tot_enrollment_q3		−0.831***	−0.663***		−2.001***	−6.341***
		-0.022	-0.053		-0.035	-0.179
ccd_tot_enrollment_q4		−1.864***	−3.424***		−2.697***	−3.249***
		-0.021	-0.055		-0.026	-0.049
URM_quartile2		−0.294***	−1.582***		−0.121***	−0.096
		-0.016	-0.055		-0.03	-0.062
URM_quartile3		−0.137***	0.013		0.102***	0.025
		-0.022	-0.057		-0.031	-0.065
URM_quartile4		0.299***	0.481***		0.496***	0.642***
		-0.027	-0.063		-0.036	-0.068
frl_quartile2		0.335***	−0.328***		0.037*	−0.833***
		-0.017	-0.058		-0.021	-0.049
frl_quartile3		0.923***	0.926***		−0.030	−0.873***
		-0.021	-0.058		-0.024	-0.054
frl_quartile4		1.037***	−0.119		0.267***	−0.306***
		-0.026	-0.075		-0.031	-0.057
asinh(count_adjust) × ccd_tot_enrollment_q2			0.095***			−0.074***
			-0.012			-0.018
asinh(count_adjust) × ccd_tot_enrollment_q3			0.002			1.040***
			-0.012			-0.04
asinh(count_adjust) × ccd_tot_enrollment_q4			0.341***			0.203***
			-0.012			-0.012
asinh(count_adjust) × URM_quartile2			0.254***			0.006
			-0.01			-0.016
asinh(count_adjust) × URM_quartile3			−0.032***			0.005
			-0.011			-0.017
asinh(count_adjust) × URM_quartile4			−0.041***			−0.091***
			-0.014			-0.019
asinh(count_adjust) × frl_quartile2			0.135***			0.211***
			-0.011			-0.01
asinh(count_adjust) × frl_quartile3			−0.023**			0.242***
			-0.012			-0.013
asinh(count_adjust) × frl_quartile4			0.282***			0.178***
			-0.016			-0.015

Num.Obs.	67442	67442	67442	41397	41397	41397
R2	0.307	0.403	0.414	0.263	0.382	0.393
R2 Adj.	0.307	0.403	0.414	0.263	0.382	0.393
AIC	202565.4	174511.1	171417.8	132329.8	110914.9	108987.8
BIC	202583.7	174647.8	171636.7	132347.1	111044.4	109195
RMSE	0.41	0.37	0.36	0.43	0.38	0.37
Std.Errors	IID	IID	IID	IID	IID	IID

Table A1D. Comparison of estimates of remote status (Nationwide)

		Probability Method[a]	Optimal Threshold[b]	P&L Method[c]	R2L[d]
Elementary Level	Probability Method	1.00			
	Optimal Threshold	0.87	1.00		
	P&L Method	0.42	0.59	1.00	
	R2L	0.19	0.26	0.45	1.00
Middle School Level		Probability Method[a]	Optimal Threshold[b]	P&L Method[c]	R2L[d]
	Probability Method	1.00			
	Optimal Threshold	0.90	1.00		
	P&L Method	0.45	0.56	1.00	
	R2L	0.12	0.19	0.42	1.00
High School Level		Probability Method[a]	Optimal Threshold[b]	P&L Method[c]	R2L[d]
	Probability Method	1.00			
	Optimal Threshold	0.92	1.00		
	P&L Method	0.48	0.53	1.00	
	R2L	0.23	0.26	0.42	1.00

a. For each school, we compute the probability a school is remote based on the logit model in Equation (1).

b. Optimal thresholds are from Table 6. We use the threshold from the associated with our method at various school levels, and with remote and in person schools weighted equally. Here, we use an optimal threshold that is the average of that from Michigan and Washington.

c. The Parolin and Lee (P&L) Method uses a year-over-year measure to determine the remote status of school. Their method is to label a school as remote in a given month if it exhibits a 50% drop-off in visits from Month X in Year Y to Month X in year Y+1.

d. R2L assumes a district operates under one modality, and scrapes data from their websites to classify whether a school is in person or remote.