School Closures, Mortality, and Human Capital: Evidence from the Universe of Closures during the 1918 Pandemic in Sweden*

Christian M. Dahl Casper W. Hansen Peter S. Jensen
Martin Karlsson Daniel Kühnle†

This version: July 26, 2024.

Abstract

We comprehensively examine the impact of primary-school closures during the 1918 influenza pandemic in Sweden on both short-run mortality and long-run outcomes of affected school children. Using the universe of death certificates and newly-collected data on school closures, we first document that schools closed due to local influenza outbreaks. To address the endogeneity of school closures, we focus on parishes that closed schools within an event-study design and document that faster school closures significantly reduced mortality rates among primary-aged individuals. Our long-run difference-in-differences estimates shows precisely estimated, minor, and mostly insignificant effects on longevity, employment, and income.

Keywords: school closures; 1918 Pandemic; mortality; human capital; short- and long-run effects

JEL Codes: I18, I10, J10, N34

*__Acknowledgments__: We would like to thank Jonas Jessen and participants at the Ninth CEPR Economic History Symposium (2022), Tinbergen Institute (Amsterdam), Western Economic Association in Portland 2022, Society of the Economics of the Household 2023 (Copenhagen), European Society of Population Economics 2023 (Belgrade), Linnaeus University (Växjö), Lund University, Gothenburg University, the Economic History Workshop at Uppsala University, the 2023 Economic History Association meeting Pittsburgh (in particular Melissa Thomasson as discussant). A research grant from the German Research Foundation DFG (Project number 467132381) is gratefully acknowledged.

†Dahl: University of Southern Denmark, cmd@sam.sdu.dk; Hansen: University of Copenhagen and CEPR, casper.worm.hansen@econ.ku.dk; Jensen: Linnaeus University, peter.sandholtjensen@lnu.se; Karlsson: University of Duisburg-Essen, martin.karlsson@uni-due.de; Kühnle: University of Duisburg-Essen, daniel.kuehnle@uni-due.de.

1 Introduction

Throughout history, major pandemics have recurrently impacted populations across the world, with the 1918-1920 Great Influenza Pandemic (henceforth, the 1918 Pandemic) ranking among the deadliest, claiming an estimated 20 to 100 million lives worldwide (e.g., Beach et al., 2022a). The pandemic unfolded in an era marked by increasing public-health awareness in several Western societies, primarily underpinned by the germ theory of disease that was developed 30-40 years prior (e.g., Alsan and Goldin, 2019; Cutler and Miller, 2005; Cutler et al., 2006). Consequently, the United States and numerous European nations instated non-pharmaceutical interventions (NPIs), including the closure of educational institutions, to contain the epidemic. However, school closures always run the risk of incurring human capital losses due to lost instruction time, which underscores a key policy trade-offs of NPIs during pandemics: weighing immediate gains against potential long-term losses.

Despite the importance of studying both the short- and long-term outcomes of NPIs, the possibilities of drawing generalisable inferences still remain impeded by the lack of suitable data. Seminal studies that have examined the efficacy of NPIs during the 1918 Pandemic in the United States (e.g. Bootsma and Ferguson, 2007; Hatchett et al., 2007; Markel et al., 2007) rely on weekly tabulated mortality data from 1918 to 1919 with limited geographic coverage of cities.[1] Additionally, the body of evidence regarding the enduring consequences of widespread school closures during major pandemics on long-term socioeconomic outcomes remains limited (Ager et al., 2022), notwithstanding a general literature on instructional time, which suggests that even brief interruptions can yield substantial ramifications on future life prospects (Cattan et al., 2022; Liu et al., 2021; Fischer et al., 2020).

In this paper, we examine the short- and long-run effects of school closures during the

[1]Newer studies like Barro (2022), Correia et al. (2022), and Berkes et al. (2020), among others, also rely on some of the same data sets. In a U.S. context, data availability creates a trade-off between coverage and frequency when using tabulated mortality data from the Vital Statistics of the United States. Moreover, complete information on schools closures is currently only available for a subset of 50 cities across the U.S. (see Influenza archive).

1918 Pandemic in Sweden, which provides us with a distinctive opportunity to study the trade-off between short- and long-term objectives of NPIs during pandemics. In particular, the possibilities to collect data on school closures are extraordinary and allow us to advance the literature along several dimensions as we can overcome many data limitations constraining previous studies. To begin with, we collected data on the timing and occurrence of school closures from local archives across Sweden for 99 percent of all school districts. Our short-run mortality analysis combines the precise dates of these school closures with the universe of individual death certificates from 1914 to 1920 (around 500,000). Our baseline sample includes information on the exact timing of school closures for around 1,300 school districts–averaging about 17 days with a maximum of 95 days–while around 830 districts maintained open schools throughout the epidemic. Furthermore, we use machine learning and handwriting recognition algorithms to transcribe the universe of death certificates for the years 1918 to 1920 (approximately 265,000) for various causes of death, including influenza, pneumonia, and tuberculosis. Among other things, we use the cause-of-death data to estimate the local arrival date of the epidemic across all areas of Sweden, allowing us to evaluate whether closing schools fast, as opposed to late, saved lives. An additional advantage of using Sweden as our laboratory is its neutrality in World War I, which meant the absence of a second mortality shock co-inciding with the pandemic.

We begin by using variation in the timing of school closures across school districts to descriptively document that schools closed, on average, around two weeks *after* local mortality rates began to rise. To address the reverse causality issue, we focus our analysis only on closing school districts and, within this group, estimate the mortality effect of closing schools fast versus slowly. For this analysis, we infer the epidemic arrival dates for all districts combining information on aggregated influenza infections with individual influenza-related deaths. This allows us to identify when districts closed their schools relative to the epidemic arrival date. Within an event study design, we then compare mortality in the

weeks before and after the closure between fast- and slow-closing districts. We estimate flat pre-trends prior to the closure, but slow-closing districts experience much higher mortality in the short run compared to fast-closing districts, demonstrating that fast school closures reduced the peak intensity of the epidemic. We estimate that closing schools fast cut the weekly epidemic mortality rate in half at the peak, but in the long run, i.e., evaluated over 20 weeks after the closure, we find the effects to be small and statically insignificant.

In general, our analysis shows that mortality differences between fast- and slow-closing districts, as well as differences between closing and non-closing districts, tend to dissipate completely within six weeks. This makes it hard to attribute the estimated effects of fast closures to differences in the natural 'cycle' of the pandemic, which lasted much longer: in our data, mortality rates return to pre-pandemic levels only after 30 weeks. Our detailed death data allows us to show that this reverting pattern is driven by old-age individuals and closing fast indeed saved lives among individuals of ages 14 to 49. To our knowledge, these patterns have not been documented before in research on the 1918 Pandemic.

Lastly, we investigate potential long-run effects of school closures in general on the affected school children, which is crucial for the cost-benefit evaluation of such policies.[2] For this part, we follow the entire cohorts of school-aged children over their lifetimes and measure their long-run socioeconomic outcomes in the census data in 1950, 1960, and 1970. For identification, we leverage the institutional rule that Swedish children started school in the late summer of the year in which they turn seven. As we know the exact day of birth of all individuals from the censuses, we can estimate the long-run effects of school closures using a difference-in-discontinuities or difference-in-differences approach. Intuitively, we compare the long-run outcomes of children who started their first year of school during the epidemic (born in 1911) with those who started school one year later (born in 1912) across

[2] A seminal contribution by Adda (2016) has highlighted that school closures may be problematic even if they reduce the spread of pandemics, if the short-term benefits are outweighed by long-term economic losses due to reduced instruction time.

closing and non-closing school districts. This identification strategy allows us to estimate the effect of closing schools in general versus keeping them open, isolating the effect of lost schooling from other effects of the school closure on mortality and other local outcomes. We show that survivorship bias is not an issue in this design. Overall, we estimate precise null effects for most outcomes, including longevity, employment, and income. Using our most precise difference-in-differences specification, we can rule out effects on longevity that are larger than 1% for men and for women, and for income around age 60 we can rule out negative effects that are larger than -0.9% for men and -0.4% for women. We find no differences between short and long closures. Thus, we conclude that the closures did not come with any substantial long-run socioeconomic cost (or benefit) for the affected cohorts.

Our paper contributes in several dimensions to our understanding of the relationship between school closures and mortality during the 1918 Pandemic. First, our study covers the vast majority of all areas in a country (Sweden) in combination with the universe of individual death certificates. This combination of coverage of school closures and high-frequency mortality data, including information on the cause of death, is to our knowledge unique in the study of the 1918 Pandemic, as the previous literature is based on U.S. data that rely on a subset of 50 cities. Second, we are the first study to show school closures during the 1918 Pandemic being implemented in response to a local surge in the epidemic, using a weekly or daily event study analysis. This pattern is less emphasised in epidemiological studies, such as Hatchett et al. (2007) and Markel et al. (2007). Our study therefore makes a methodological contribution that cautions against comparing mortality rates between closing and non-closing areas without having detailed information on pre-closure epidemic trajectories. Third, our ability to identify the local epidemic arrival date allows us to show that closing schools fast reduced the intensity of the epidemic and saved lives. This result fits well with previous descriptive studies for the U.S., such as Hatchett et al. (2007), Markel et al. (2007), (Barro, 2022), and Correia et al. (2022) who arrive at similar conclusions in

terms of implementing NPIs early. Our paper contributes to this research by using more granular data, which allow us to assess pre-trends at the weekly/daily level, applying a causal identification strategy, and breaking down the findings by age to document possible short-run inequality implications of the closures.

Our study on one of the most deadly pandemic in modern history is of course related to the literature on the Covid-19 pandemic, despite the striking differences between the policy environments and contexts. As the Covid pandemic has unfolded in a much more developed environment, governments throughout the world implemented various NPIs and researchers collected these data and studied their effectiveness almost in real time (among many others, see Brauner et al., 2021; Chinazzi et al., 2020; Cho, 2020; Dehning et al., 2020; Elenev et al., 2020; Flaxman et al., 2020; Haug et al., 2020; Hsiang et al., 2020). These two pandemics, however, have been quite different in terms of age-specific mortality rates, where the 1918 Pandemic is famously known to be W-shaped and Covid-19 monotonically increasing in age. Since the 1918 Pandemic unfolded more than 100 years ago, our study is able to study the long-run effects of school closures on socioeconomic outcomes. Overall, our results indicate null effects which are precisely estimated. Although we cannot reject that school closures had no long-run impact on different socioeconomic outcomes, we can rule out any substantial long-run negative (or positive) effects on the school children they affected.

Only one paper has so far analysed the long-run effects on education and earnings of school closures during the 1918 Pandemic, and this is for the US. Ager et al. (2022) report precise null effects of school closures on adult educational attainment, wage and non-wage income, and hours worked in 1940 for a sample of males living in 168 large cities linked across the 1920 and 1940 censuses. Our study differs from theirs in several important respects. First, we are able to carefully document the endogeneity of school closures, and find that the issue of reverse causation seriously limits the possibilities of conducting causal inference. Nevertheless, since our analysis is based on the universe of the Swedish population and uses

historical individual-level data, we are able to develop two different causal identification strategies–one for the short-run, one for the long-run–that deal with the endogeneity of school closures. We show that there are no systematic differences in mortality rates between treated and untreated children ('survivorship bias') that could affect our long-run estimates. Moreover, having not only males in urban areas but the entire population of males and females across all school districts of the country enables us to generate estimates for the entire population and also study effect heterogeneity along important dimensions. The latter point is important considering consistent evidence that males and females respond differently to extensions to compulsory schooling (Cygan-Rehm, 2022; Fischer et al., 2022, 2020). We are furthermore able to study effects on additional outcomes (e.g., longevity) using death certificates and measure the individual socio-economic outcomes at multiple points in time over their life cycle until 1970. In addition, we can use individual student absence data during the pandemic to show that school closures actually reduced student absences. Despite these differences, our more comprehensive analysis also demonstrates that the relatively short school closures during the 1918 Pandemic did not adversely impact long-term socioeconomic outcomes, while providing new evidence for short-run mortality.

Our work is also related to studies investigating different aspects of the 1918-Pandemic from its effects on the economy (e.g., Barro et al., 2020; Correia et al., 2022; Dahl et al., 2022; Karlsson et al., 2014; Velde, 2022) to effects on human capital (e.g., Almond, 2006; Beach et al., 2022b), fertility (Boberg-Fazlic et al., 2021), the health-care sector (Esteves et al., 2022), and the interaction between political rule and innovation (Berkes et al., 2020; Xu, 2021). Overviews of the current literature are provided by Beach et al. (2022a) and Karlsson et al. (2022). More broadly, our paper also contributes to the returns to schooling literature (e.g., Card, 2001; Jensen, 2010; Rosenzweig, 1995), in particular to studies examining the effects of instructional time on student outcomes (Aucejo and Romano, 2016; Cattan et al., 2022; Goodman, 2014).

2 Background

As in most other parts of the world, the 1918 Pandemic reached Sweden–then with a population of 5.8 million–in three major waves, with two waves in 1918 and one subsequent wave during the spring of 1919. The first cases in Sweden were recorded in the cities of Malmö and Gothenburg in June 1918; this first wave was brought to Sweden by migrant workers returning home to celebrate Midsummer. Initially, the seemingly mild flu caused little concern. During the first seven months of 1918, 148 influenza deaths were reported, which is below the corresponding figure for 1917 (190 influenza deaths, see Karlsson et al., 2014). The second wave, which was characterised by particularly high mortality rates, entered the country in the North, via the railway line connecting the Norwegian city of Trondheim with the Swedish city of Östersund (Åman, 1990; Mamelund, 1998). Thus, there was a distinct upsurge in mortality noted from late August 1918 onward, and the second wave spread rapidly throughout the country with a notable spike during October and November. The second wave was responsible for the majority (around 70 percent) of the 38,500 deaths; 20 percent of the deaths occurred during the spring wave of 1919, whereas a fourth (more limited) wave was responsible for less than 10 percent of the deaths, and this only affected some parts of the country (National Board of Health, 1920, 1921, 1922).

The death toll of the pandemic exhibited substantial regional variation within the country, with some areas experiencing up to three times higher rates than others. The northern counties Jämtland, Norrbotten, and Västernorrland were particularly severely hit. The high mortality rates in the northern areas have, in part, a demographic explanation as these regions had a young population. However, it has also been hypothesized that the high regional variation in mortality rates may be explained by remoteness, and that people living in these areas had less immunological protection against the virus as they had been less exposed to earlier flu waves. Karlsson et al. (2014) argue that regional differences in excess

mortality are largely unrelated to observable characteristics such as population density, earnings, and the sectoral composition of the economy. They also note that several key economic indicators develop in parallel across more and less hard hit regions during the years preceding the pandemic. Karlsson et al. (2021) study excess mortality at the parish level and note that it correlates strongly with a number of variables, including population density, infant mortality in previous years, access to the railway network, and weather in September. They also report that the correlations appear to be different in urban and rural areas, sometimes even with opposing signs.

Authorities at all levels of government responded in different ways to the public health crisis brought by the pandemic. The national government made attempts to mobilize health-care staff to the most hard hit areas in the north, with limited success. Some cities also mobilized medical resources at the local level by reallocating personnel responsible for tuberculosis prevention and for investigating sanitary conditions in homes to instead combat the pandemic. The available resources were, however, inadequate everywhere, and this also holds for the "epidemic hospitals" that cities were obliged to operate since 1874, as they were not planned for epidemics of this scale (Åman, 1990).

Instead, NPIs became the most important measure in the struggle to contain the pandemic. These NPIs were of two different kinds: school closures and public gathering bans. In the cities, the public health authority (*Hälsovårdsnämnden*) was responsible for taking action; these actions had typically been initiated by the city physician. In rural areas, the local councils were responsible for the prevention and mitigation of epidemics (Åman, 1990). If they concluded a public gathering ban was necessary, they would forward a request to the regional governor (*länsstyrelsen*) who would issue an announcement to be read out in churches in the affected parish. School closures were typically initiated by the chief medical officer of a health district, which typically consisted of a handful of rural parishes. The school closures would typically apply to all the schools in a school district; however, in some

cases, only individual schools or even classes were closed, whereas the others remained in normal operation.

To separate the potential impact of gathering bans on mortality from the impact of school closures, we also collected data on gathering bans from two sources: First, the announcements of the county governor, which mainly cover the rural parishes. For urban parishes, we checked the protocols of individual city public-health authorities (as provided by local archives) and supplemented these with information from newspapers of the time and various other national and local sources. We end up covering about 30 percent of all parishes with information on gathering bans and show empirically that these do not confound our school-closure findings.

3 Data, sample, and descriptive patterns

3.1 School closures

The possibility to collect data on school closures in Sweden are extraordinary, since detailed information is available for the universe of schools and pupils from two alternative sources. The first source is the 'diary with exam catalog' (*Dagbok med examenskatalog*; "exam catalog" henceforth), an administrative source which is kept in each of the 290 municipal archives in Sweden.[3] The standardised form, which was used in all schools in Sweden throughout the six years of primary school, was introduced in the academic year 1917-18, though it had precursors that contained more or less the same information. The exam catalogs included a diary, in which the daily absence or presence of each individual pupil was recorded throughout the academic year. It also had a separate table recording dates of school closures by reason. In the vast majority of cases, the exam catalog provides the exact dates of school

[3]For a more detailed description of the information included in the exam catalogs see Bhalotra et al. (2021); Cattan et al. (2022); Fischer et al. (2021).

closures due to the influenza pandemic. In rare cases, the dates are not provided in the separate table, but can be inferred from notes in the diary. Appendix Figures B.1 and B.2 provide extracts from an exam catalog.

The data collection was carried out by sending instructions to all 290 municipal archives. The archives were instructed to send copies of one exam catalog from the largest school in each district, and of one catalog from another school.[4] In case there was only one school in the district, the archive was asked to send one copy from lower primary school (grades 1-2) and one copy from upper primary school (grades 3-6). The reason why two copies were ordered for each district was to be able to assess the heterogeneity within the school district. For each of the 2,430 school districts, the school closure (or non-closure) applying in grade 3-6 in the largest school was taken to represent the entire district. In case the other school sampled from the district had conflicting information, it was recorded as a separate variable. In case the same closure dates applied also in the second school, it was assumed that it applied uniformly in the entire district.

A second source containing the same information in a different format are the teaching statistics forms, which were collected by the national ministry of ecclesiastic affairs since 1915. Each school in the country had to complete the form every year. One copy of the form is available at regional archives in the collections of each former school district, and one copy is available at the national archive in the collection of the ministry of ecclesiastic affairs. We digitised this source in its entirety and used it to complement and validate the main dataset. An example of this source is provided in Appendix Figures B.3 and B.4.

Together, the two sources provide information on school closures during the pandemic for 99 percent of the 2,430 school districts. We ascertain the exact closure and reopening

[4] In 1918 Sweden consisted of 2,430 school districts. In almost all cases, school districts corresponded to a municipality, of which there were 2,511 in the same year. Some smaller rural municipalities would form a joint school district consisting of 2-4 municipalities. Municipalities were of three types: market towns (*köpingar*), cities (*städer*) and rural parishes (*landskommuner*). Population records were kept by the church and thus place of residence was recorded as a church parish: there were in total 2,587 church parishes which in general match perfectly into municipalities and school districts (Statistics Sweden, 1918).

dates for around 1,300 districts, with a mean school closure duration of 17.8 days and an interquartile range of 10 to 24 days (maximum of 95 days). Additionally, around 830 districts maintained open schools throughout the epidemic. In Panel A of Appendix Figure A.1 we plot the spatial distribution of closure dates across districts, showing that school closures were widespread across the country and not concentrated in specific regions.

To assess the validity of the reported dates, we performed two exercises. First, we cross-checked the two sources by sampling districts from five different regions and found that only two out of 485 had conflicting information between the two sources. Further examination revealed that these discrepancies were likely due to the way school closure days were reported, as some statistics may have netted out days that were later retaken. Second, we additionally digitised and linked a subset of exam catalogues to censuses; the sample consists of around 13,000 individuals from 362 school districts. For this sample, we then examined how the length of school closures mapped into days of lost schooling. Our analysis based on this supplementary sample suggests that one day of school closures reduced the number of days spent in school per year by approximately 0.45 days. We discuss the reasons for these deviations at length in Online Appendix D. Nevertheless, this supplementary analysis clearly documents a strong negative relationship between closures and presence in school, hence confirming the validity of our school closures data.

3.2 Mortality

We generate our mortality data from the complete set of death certificates issued between 1914 and 1920, comprising over 500,000 records from the Federation of Swedish Genealogical Societies (2018). These certificates contain various details, such as the exact dates of death and birth; parishes of death; and gender. We use this information to calculate the number of deaths occurring in each parish on a weekly or daily basis. Importantly, the geographical information in the certificates refers to the parish of residence, rather than the place of

occurrence. This detailed geographical information better suits our analysis and offers an additional advantage compared to historical U.S. studies, where tabulated death counts are often based on place of occurrence.[5]

We calculate the parish-level weekly (or daily) mortality rates by combining death counts from the certificates with annual parish-level population data. These population figures represent the number of people at risk at the beginning of each year, and we do not interpolate within a year.[6] We also calculate age-group specific mortality rates for each parish taking the number of people at risk for each age group in each parish from the 1910 full population census.

Furthermore, we employ machine learning and handwriting recognition algorithms to digitise the cause-of-death from around 265,000 death certificates between 1918 and 1920. Our cause-of-death data includes information on deaths related to, e.g., influenza, pneumonia, tuberculosis, cancer, old-age, and heart diseases. Online Appendix C provides detailed information about our transcription methodology.

To visualise the spatial pattern of excess mortality rates, we calculate the difference between the mortality rate during the epidemic period (June 1918 to March 1919) and the average mortality rate during the same months in the previous four years (1914-1917), expressed as the number of deaths per 1,000 people in each parish. Panel B of Appendix Figure A.1 illustrates the absence of pronounced geographical clusters in terms of the local burden of the epidemic, similar to the findings for school closures.

[5] Mortality rates based on place of occurrence can, for example, misrepresent areas with epidemic hospitals as more deadly than they actually are.

[6] Our results remain robust under alternative assumptions regarding when to measure the population at risk. For instance, we obtain consistent results throughout our study when fixing the population to pre-epidemic levels, as measured in January 1918.

3.3 Identifying fast and slow-closing school districts

In our analysis, we aim to classify school districts into fast and slow school-closing groups. This classification requires information on local epidemic arrival dates for each district, which is not available in the archives. Consequently, we combine several data sources to infer these local epidemic arrival dates, which enable us to determine the timing of school closures relative to the beginning of the local influenza outbreak.

First, we use data on the number of influenza infections at the monthly level, which is unavailable at the local level but can be obtained for health districts from 1916 onwards.[7] With this information, we calculate the excess influenza infection rate for each health district, measured as the monthly deviation from the average infection rate for the period between 1916 and 1921, excluding the pandemic years of 1918 and 1919. Appendix Figure A.2 demonstrates that excess infections begin to increase in July 1918, peak in October 1918, and return to pre-pandemic levels in July 1919, followed by a final (but less deadly) wave in the first quarter of 1920.

Second, we use this information to identify the local epidemic arrival date as the month when the district-specific excess infection rate starts to increase substantially.[8] We then estimate event study models of the local epidemic arrival on excess influenza rates, controlling for district-month effects. Panel A of Appendix Figure A.3 shows that our method works well in estimating the month of arrival.

After determining the local month of arrival, we employ information from the weekly cause-of-death data to identify when confirmed influenza deaths occur within (or after) the inferred arrival month. We use information on the exact timing of the first four weeks with confirmed influenza deaths to establish the local week of arrival. To account for geographic clustering, we calculate the median arrival week at the city or district (*härad*) level. From

[7] There were approximately 450 health districts in Sweden in 1918, with each district containing an average of 5 school districts.

[8] For this, we set the cut-off for the excess influenza rate per 1,000 population at 0.5.

this date, we subtract two weeks to account for a median time from incubation to death of 7–11 days (Klugman et al., 2009). We validate our approach by estimating event study models using the Gardner (2022) estimator to account for dynamic treatment effects. Panel B of Appendix Figure A.3 shows how parish-level mortality and influenza rates begin to increase around the estimated arrival date and return to its original level within 30 weeks.

Next, we illustrate the timing of the inferred weekly epidemic arrival dates against the timing of the school closure dates in Panel A of Appendix Figure A.4. In Panel B, we represent the distribution of the response time, which captures the difference (in weeks) between the school closure and the inferred arrival of influenza. The mean response time corresponds to two weeks, with a median of three weeks. We classify districts as closing faster if they close within two weeks of the epidemic arrival date. Based on this distinction, we identify 548 fast-closing districts, 546 slow-closing districts, and 828 non-closing districts with an estimated epidemic arrival date.[9]

3.4 Descriptive statistics

Table 1 displays summary statistics for various variables, grouped by districts that closed schools and those that kept them open.[10] We report group differences in three ways: First, the raw difference corresponding to a simple t-test; second, an adjusted group-difference where we control for whether a city was present in the area and the area population size in 1917; and finally, the standardised difference in the final column. Panel A presents pre-epidemic socio-economic characteristics, such as population size, tax revenues, and expenditures. Panel B covers national vote shares from the last pre-epidemic national election in 1917,

[9]We also show that our findings for fast vs. slow closures do not hinge on this classification, as we vary the threshold for closing fast and find the same results, see Appendix Figure A.15.

[10]Table 1 reports information for closing districts with exact dates, as we can leverage the timing of closure dates only for these districts in our event study analysis. In Appendix Table A.1, we compare closing districts with known and unknown closure dates, and show that the excluded closing districts (N=299) are less urban compared to districts with known closure dates.

and Panel C features pre-epidemic healthcare characteristics. The statistics reveal that districts closing their schools tend to be more urban, having larger populations and a higher likelihood of being classified as a city. This urban-rural divide is also evident in many other pre-epidemic outcomes, including per capita measures of factors like workers, factories, train stations, income, physicians, and nurses. The political dimension reflects this as well, with a higher share of the vote going to the labour party in these more urban districts that were more likely to close schools.[11] Many of the raw differences disappear or decrease substantially in size and statistical significance once we control for city status and population size, indicating that differences in urbanity drive many of the observed differences.

Figure 1 shows the average weekly mortality rates by school closure in calendar time from May 1918 and one year forward, for four different measures of mortality: overall (Panel A), excess (Panel B), influenza (Panel C), and all-other causes mortality (Panel D).[12] We see that mortality rates in closing and non-closing school districts followed a similar trajectory until early autumn, after which the all-cause, excess, and influenza-mortality rates began to diverge between the two groups. The closure group experienced a peak in all-cause mortality rate of around 0.8 deaths per 1,000 people per week in November 1918, see Panel A. After November 1918, the mortality rates in all three categories decreased and fell below those of the non-closure group for the closing districts. Although the mortality rates converged fairly quickly in early 1919, we also show in Appendix Figure A.5 that the cumulative mortality rates remained permanently higher for closing compared to non-closing school for all measures of mortality. Finally, Panel D of Figure 1 shows minor differences between closure and non-closure districts in all other causes of death (than influenza) across the

[11]While our high-frequency event studies allow us to control for week/day and parish fixed effects and thus hold any pre-epidemic level difference constant, all our conclusions remain robust when controlling for these local characteristics interacted with week/day fixed effects.

[12]We apply a three-week moving average to these mortality rates to reduce weekly fluctuations. The excess mortality rate was calculated as the difference between the weekly mortality rate in 1918 and the average mortality rate for the corresponding week over the pre-epidemic years from 1915 to 1917. The all-other causes category includes all causes of death except for influenza-related deaths.

pandemic year, which is also reflected in the small differences in cumulative mortality for this outcome.

4 School closures and mortality

4.1 Empirical strategy

As a starting point, we examine the changes in mortality during the weeks when school closures were implemented. In the literature studying the effects of NPIs during a pandemic, it has been common to exploit variation in timing of policies in a difference-in-differences framework (Haber et al., 2022). However, it has been noted that the highly non-linear behavior of incidence and mortality rates in a pandemic makes inference based on a common time trend assumption challenging (Callaway and Li, 2023). In order to assess the extent of the issue in our application, we study the evolution of weekly mortality rates in school districts before and after school closures compared to districts without school closures conditional on parish and calendar week fixed effects. The estimation equation takes the following form:

$$m_{it} = \lambda_i + \delta_t + \sum_{j=-20}^{T} \beta_j \times \mathbb{I}\left[t - S_i = j\right] + \mathbf{Z}'_{it}\mathbf{\Phi} + \varepsilon_{it}, \tag{1}$$

where m_{it} represents the mortality rate in parish i during week t (i.e., the weekly number of deaths scaled by the parish population on January 1st 1918). Parish and week-date fixed effects correspond to λ_i and δ_t, respectively. The week-date of the first school closure in parish i is S_i, while $\mathbb{I}\left[t - S_i = j\right]$ is an indicator function for being j weeks from the school closure. The omitted comparison is 20 weeks before ($j = -20$). The main coefficients of interest, β_j, trace out the dynamic development of mortality rates before and after the school closures net of parish and week-date fixed effects. In the robustness analysis, the control

vector $\mathbf{Z}'_{it}$ includes pre-epidemic parish-level characteristics interacted with week-date fixed effects (see details below) as well as county-by-week fixed effects. The error term, clustered at the parish level, is denoted by ε_{it}.

Our event study window spans 20 weeks before the school closure ($j = -20$) and 20 weeks after ($j = 20$), with event time outside the event window binned at the beginning or end of the window for the treated school districts. The baseline sample focuses on all school closures that happened during 1918, with the first school closure occurring on the 17th of July 1918.[13] Our base sample includes 2,126 districts, consisting of 1,298 closing districts and 828 never-treated districts. Additionally, we also present estimates from a daily event study, where δ_t signifies day fixed effects and the event window extends from 40 days before to 40 days after the school closure.

4.2 Baseline results

Panel A of Figure 2 presents the baseline TWFE event study estimates. The dashed vertical line separates time before and after the school closure. The black coefficients correspond to the actual pandemic starting in 1918, while the grey placebo estimates are derived using the previous year. The figure shows that closing and non-closing school districts were on similar weekly mortality trends up until two weeks before the school closure. However, we then observe that mortality rates start to increase rapidly in the closing districts two weeks before the closure. This suggests that schools closed, on average, in response to a local surge in the epidemic. Mortality rates peaked one week after the closure, with the coefficients indicating an additional 0.4 deaths per 1,000 people in closing districts relative to non-closing districts. This magnitude corresponds to around 1,540 additional deaths per week, using the total population size of the closing areas (see Table 1). Six weeks after the

[13] This captures the vast majority of school closures, as only 17 districts closed their schools in 1919 compared to roughly 1300 closures in 1918.

closure, the mortality rate returned to levels observed in non-closure districts.

Panel B of Figure 2 displays the daily event study estimates, where the week-date fixed effects have been substituted with day fixed effects and the event-window ranges from 40 days before to 40 days after the closure. The same overall conclusion emerges, but we gain further insight into the dynamics around the closure: the daily estimates show that mortality rates begin to rise 12 days before the closure, reach a peak 9-10 days after, and then gradually return to levels observed in non-closing districts around 40 days after the closure.

Our robustness checks reveal that we do not observe a similar pattern in the placebo periods, for which the event study coefficients are consistently close to zero and statistically insignificant throughout (see Appendix Figure A.6 in addition to the just reported grey estimates). These findings indicate that the pattern observed in our baseline estimates is unique to the pandemic year of 1918 and not, for example, explained by any seasonal mortality effects that happens to be correlated with school closures. We also adjust the baseline estimates using population weights (see Appendix Figure A.7) and additionally control for gathering bans (see Appendix Figure A.8), but neither change affects our main estimates.[14] Appendix Figure A.9 also shows that we get the same results using the recently developed estimators by Sun and Abraham (2020), Gardner (2022), Callaway and Sant'Anna (2021), and Borusyak et al. (2021).

Next, in Figure 3, we report the relationship between school closures and excess mortality (Panel A), influenza-related deaths (Panel B), old-age (Panel C), tuberculosis (Panel D), cancer (Panel D), and all non-influenza mortality (Panel E). We find that the estimates for excess mortality rates are very similar to those for the all-cause mortality rate. Comparing Panels A and B also shows that the majority of the excess deaths are caused by influenza deaths. This result is consistent with the other outcomes as we find only small, if any, effects on old-age

[14]The baseline findings are also robust to controlling for all the unbalanced pre-epidemic characteristics, reported in Table 1, by week-date fixed effects. Results available upon request.

mortality, nothing on tuberculosis and cancer, and small increases in the residual category of all non-influenza deaths. This detailed analysis by cause of death therefore supports the conclusion that schools were closed in response to a local influenza epidemic.

Figure 4 presents our findings for age-group specific mortality rates.[15] Our analysis reveals that the well-known W-shaped age pattern of the 1918 Pandemic carries over to school closures as well, with high mortality rates observed among young children (Panel A), prime-aged adults (Panel D), and older people (Panel E and F). Notably, we find no evidence of increased mortality rates among primary-school-aged children prior to the closure (Panel B). Our results therefore indicate that schools were closed primarily due to increased mortality rates among individuals outside the primary school environment.[16]

4.3 Estimating the effect of fast closures

So far, we have shown that school districts closed schools in response to a rise in influenza-related mortality during the two weeks preceding the closure. This reverse causality pattern prevents us from giving the estimates a causal interpretation, and we are unable to find suitable control parishes for closing parishes in terms of pre-closure mortality within a matching design.[17] To investigate the causal effect of school closures on mortality further,

[15] For these regressions, we use the population size of the same age groups from the 1910 census to proxy the population at risk. We also winsorize mortality rates to reduce the effect of outliers that render these estimates rather imprecise.

[16] As an alternative identification strategy, we also exploit variation in term-start dates within an event study design to estimate the effect for school *openings*—after the summer-break—on mortality. These opening dates were determined prior to the pandemic at the local level, with the vast majority of openings occurring between August 1st and September 16th (around 90%). In this design, everyone is treated within a period of seven weeks, and there is no never-treated group, implying that only short-term effects for up to five weeks after the school openings can be estimated. We apply the Borusyak et al. (2021) estimator and plot the event study coefficients in Figure A.11. We find flat pre-trends and cannot reject no effect of school openings on mortality in the following five weeks. We do not, however, apply this research design to term-end dates as there is little variation these, such that event time become highly collinear with calendar time.

[17] Specifically, we matched on various pre-closure characteristics, including daily mortality rates leading up to the school closure. We obtained parallel pre-closure trends, but still found positive mortality effects in the four weeks after the closures. Thus, even matching on a rich set of observable characteristics fails to account for reverse causality. For additional details and estimates, see Appendix Figure A.13.

we modify the previous empirical framework by focusing only on closing districts and by additionally exploiting the response time of districts. Intuitively, we aim to compare fast- and slow-closing districts in a manner that approximates random variation in closure response time.

For this part, we employ our classification of school districts into fast- and slow-closing groups (see Section 3.3) and begin by assessing the balance of socioeconomic characteristics between the two groups of districts. Table 2 shows that fast-closing districts were less likely to be in cities, had smaller populations and generally fewer attributes linked to urban areas, such as workers, factories, train stations, hospitals, etc. Many of the standardised differences are larger than 0.2, suggesting relative large differences in the raw comparisons. However, most of these differences become smaller in magnitude and statistically insignificant once we control for the presence of a city and populations size. When we exclude cities and districts with more than 5,000 residents, which we refer to as the "rural sample", Appendix Table A.2 reveals that the remaining fast and slow-closing districts balance on most socioeconomic pre-epidemic characteristics, apart from population size, which correlates significantly with response time (see also Appendix Figure A.12). With controls, we find that almost all significant differences disappear, with the exception of nurse density. Apart from population size and nurse density, the standardised differences are all below 0.2, and the majority even below 0.1, which represents an order of magnitude common even in randomised experiments (Imbens and Rubin, 2015, p. 352). This finding mitigates concerns about systematic differences in socioeconomic traits between fast- and slow-closing districts. In the subsequent analysis, we focus on this more balanced sample of rural districts, although the results also hold when we include all districts as in Table 2.

We first examine the raw mortality trends over time based on response time—fast, slow, never. Figure 5 demonstrates descriptively that, prior to the pandemic, all three groups follow very similar trends. The groups start diverging in mid-September 1918 with closing

school districts experiencing higher mortality compared to non-closing districts. For a while, fast-and slow-closing districts follow similar trajectories, but, starting in October 1918, fast-closing districts experienced lower mortality rates during the epidemic compared to slow-closing districts, as indicated by Panels A to C. Appendix Figure A.14 reveals a distinct ordering concerning cumulative mortality: Starting from identical trends, the school-closing districts experience a more rapid increase in excess mortality as the pandemic sets in compared to never-closers. The trends between fast- and slow-closures also coincide for some time. However, the trends subsequently diverge, and slow-closures undergo the most significant rise in excess mortality. According to Panel B in the Appendix Figure A.14, during the Spring of 1919, the accumulated excess mortality rate remained constant in all three groups of parishes, but with 0.4 fewer deaths per 1,000 people in the fast-closing districts compared to the slow-closing districts.

We next examine this pattern in our event study framework, running the analysis separately for fast- and slow-closing school districts compared to non-closing districts. Panel A of Figure 6 presents the event-time coefficients. As expected, the figure shows that mortality begins to rise earlier in slow-closing districts compared to fast-closing districts, resulting in significantly higher peak mortality rates for slow closers during the three weeks following the closure. However, the two groups converge rather quickly again, with slow-closing districts returning to similar levels of fast-closing districts four weeks after the closure.

We can exploit this pattern to arrive at causal estimates of fast school closures by shifting the event-time for slow-closing school districts one week forward, see Panel B of Figure 6, making $t = 1$ correspond to the closure week for slow-closing districts in this panel. In Panel B, the dashed lines indicate the pre-closure period for each group, the solid lines represent the post-closure period. This adjustment more clearly exposes the epidemic dynamics: both types of districts experienced an influenza outbreak with a similar trajectory, but fast-closing

districts responded more rapidly to the same mortality increase by closing schools earlier compared to slow-closing districts. Consequently, fast-closing districts managed to "flatten the curve" compared to slow-closing districts, which experienced higher peak mortality rates. For example, three weeks after, the estimated effect for the fast closures is almost half the size of the slow closures. However, they return to the same level as fast-closing districts within a period of four weeks.

To estimate the effect of the fast vs. slow closures, we run the event study analysis in two versions: First, only for closing school districts where we interact the event-time dummies with a fast-closure indicator variable. Second, we also include the never-closing districts to identify the time-effects within a triple differences design (Olden and Møen, 2022). Panels C and D of Figure 6 plot the interaction effects confirming that fast- and slow-closing districts followed identical mortality trends until the fast districts closed their schools. This finding supports the parallel trends assumption required for a causal interpretation of these estimates. Mortality then begins to decline in the first week after the fast school closure, which aligns with the time to death, and remains significantly lower for fast-closing districts compared to slow-closing districts for approximately four weeks. The magnitude of the effect is substantial: $\hat{\beta}_3 \approx -0.28$ suggests that a fast school closure resulted in 0.28 fewer deaths per 1,000 people in the third week after the closure, corresponding to around 211 lives saved during that week (multiplying the coefficient with the population-at-risk, i.e., $0.28/1000 * 530 * 1421.35$). The joint effect from week 1 to 6 after the fast closure is -0.91 ($p < .01$), indicating that around 686 lives were saved cumulatively during those six weeks; the combined long-run effect over the first 20 weeks is small and statistically insignificant in both designs (e.g., 0.03 in the DD design with $p = 0.93$). These findings align with the descriptive patterns for cumulative mortality (see Panel A of Figure 5) and previous evidence for the U.S. (Barro, 2022), who also finds that faster NPIs flattened the mortality curve in the short run, but the overall effect on mortality is small and insignificant.

Appendix Figure A.15 shows that our results are robust to using alternative cut-off values to be a faster closer. In particular, here we vary the cut-off threshold from three weeks before to four weeks after the epidemic arrived. The results demonstrate similar effects throughout all specifications, suggesting that it always saved lives to closer faster rather than later. This implies a monotonic dose-response in closing fast.

To explore heterogeneity in the effect across age groups, we demonstrate in Figure 7 that the small positive, though insignificant, longer-run effect is driven by older-age mortality rates. In particular, we observe that closing fast significantly reduced the long-run age specific mortality rates for individuals of ages: 14-19 (Panel C), and 20-49 (Panel D), while there is a reverting pattern for people of older ages (Panel F) in which closing fast reduced mortality rates the first three weeks after the closure, but in the later weeks the fast-closing school districts experienced higher mortality rates among these ages. This suggests that more frail individuals were protected in the short run due to the fast closure, but not in the long run.

5 Long-run effects of school closures

In this section, we examine how the school closures affected long-run outcomes of the affected school children up to 50 years after the closures. For the long-run outcomes, we additionally use information from the 1950, 1960, and 1970s censuses, as well as death certificates, which allow us to measure the length of life for all individuals. Both the 1950 census and the certificates provide data regarding the parish of birth, which typically aligns with school districts (as explained in footnote 4), in addition to individuals' ages. With this information, we can determine whether individuals underwent a school closure during the 1918 Pandemic or not. Moreover, the 1950 census facilitates a linkage to subsequent censuses through distinctive identification numbers, thereby enabling the tracing of the

same individuals across multiple census waves.

The census data contains information at the individual level on educational achievement, employment, income, and occupational information. All censuses contain information on whether someone worked during census week or the entire calendar year. Only the 1950 census contains the HISCAM score for the type of job performed (i.e., an occupational-based income score). For educational attainment and socioeconomic outcomes, we use information from the 1960 and 1970 censuses. From the 1960s census, we have information on educational attainment, specifically whether an individual finished "gymnasium" (equivalent to A-levels), granting access to higher education. The 1970 census contains information on earnings and, for people born after 1910, employment status and additional information on whether someone completed secondary education.

To identify the causal effect of school closures on long-run outcomes, we additionally exploit the institutional rules that children in Sweden, at the time of the pandemic, started school in the year they turned seven. This introduces a sharp discontinuity in school starting age on January 1st of each year. As the school year begins in fall of each year, children born on December 31st 1911 entered school in the fall of 1918, whereas children born one day later (on January 1st 1912) entered school in the fall of 1919, i.e. after the pandemic. Thus, we can compare the difference in outcomes between children born before and after the school entrance cut-off in parishes with/without school closures either within a difference-in-differences or difference-in-discontinuities setup. As this is not a staggered roll-out design, we do not need to apply the recently developed estimators (Roth et al., forthcoming). Further, we can control for parish fixed effects and thus condition on local conditions, in particular pandemic severity. Note that this design thus nets out any potential school-starting age effects and any age-specific mortality effect from the pandemic itself. Accordingly, the empirical strategy in the long-run analysis isolates the effect from being

absent from school due to the closure from the pandemic itself.[18]

For our baseline analysis, we estimate the following difference-in-differences model:

$$y_{impy} = \gamma_p + \gamma_m + \gamma_y + \gamma_D D_i + \tau C_p \cdot D_i + \epsilon_i, \tag{2}$$

where y_{impy} represents the outcome of person i, born in parish p in month m and year y. The parameter γ_p denotes a vector of parish fixed effects, which subsumes a time-invariant treatment indicator for each parish ($treat_p$); γ_m and γ_y are fixed effects for birth months and birth years. The indicator variable D_i indicates whether a child was in school during the pandemic, and the indicator variable C_p is equal to 1 if a district closed schools during the pandemic. Thus, the coefficient τ identifies the difference-in-differences estimate for the effect of school closures.[19] For the baseline specification, we include children born within a two-year window around the relevant school-entrance cut-off, i.e. born in 1910–1913. We later show that our baseline results are robust to the choice of this observation window.

We also estimate a difference-in-discontinuties model of the following form:

$$y_{icp} = \alpha_p + \alpha_1 r_i + \alpha_2 D_c + \alpha_3 C_p \cdot D_c \beta_1 C_p \cdot r_i + \beta_2 D_c \cdot r_i + \beta_3 C_p \cdot r_i \cdot D_c + \epsilon_i,$$

where y_{icpt} denotes the outcome of individual i, born in year c in parish p. The parameter α_p represents a vector of parish fixed effects, which again subsumes the treatment indicator for each parish (C_p). We include the running variable, i.e. children's birth date r_i as a linear term; for simplicity, we recenter the date of birth so that the January 1st cut-off corresponds

[18]We focus on school closures and not on fast versus slow closures, as the speed of closures should not matter in the long-run.

[19]In our long-run analysis, treatment assignment is based on parish of *birth*, assuming children went to school where they were born, which is a common supposition in the literature (e.g., Ager et al., 2022). However, our collected datasets provide a unique opportunity of a critical assessment of this matter. Specifically, we demonstrate in Appendix D.4 that the vast majority individuals, in fact, went to school in their parish of birth. But we also estimate the bias arising from implementing the DID specification with a mismeasured treatment exposure and show that the measurement error in the treatment assignment based on parish of birth leads to a downward bias of estimates by around 26 percent.

to $r = 0$. We define the indicator variable D_c as $D_c = I(r_c > 0)$. We fully interact all terms, so that the interaction term α_3 identifies the difference-in-discontinuities estimate of school closures during the pandemic on the long-run outcomes of individuals. As in the difference-in-differences models, we vary the observation window from 12 months to 6 months on either side and show that the results are robust to this modelling choice.

To get a visual impression of the trends in long-run outcomes, we plot the average values of several long-run outcomes (employment, income, retirement, and longevity) by birth cohorts, sex and treatment status (i.e., whether schools were closed or not) in Figure 8. This graphical evidence gives a first indication that outcomes, in general, develop quite smoothly by birth cohorts and do not exhibit any large discontinuities, for example, between the cohort just too young to be in school during epidemic (control) and the one-year older cohort (treatment) attending school during that year (treatment). Due to the substantial differences in the educational and labour market outcomes by sex, we perform the regression analysis separately for both sexes.

We report the regression results from the DiD and DiDisc models in Table 3. We show the results for men in columns 1 and 2 and for women in columns 3 and 4. Panel A starts with the mortality outcomes. First, we show that closure is unrelated to a placebo outcome (dying before age 5, which is one year earlier than the younger discontinuity cohort). Second, we find no difference in the probability of dying during the pandemic for children experiencing the closures relative to the control children. Third, we exploit that we have information on the length for life for all children attending school during those years and find no effect on the length of life. This effect is statistically insignificant and economically small as we can rule out effects larger than -1% and +0.3% for men, and -.9% and +0.4% for women respectively, using the 95% confidence intervals in the DiD specification.[20]

[20]For instance, for men the mean age at death for those born between 1910 and 1914 is 68.9, and the 95% confidence DiD interval ranges from -0.692 to 0.188. For women, the corresponding mean age at death is 68.5 and the the 95% confidence DiD interval ranges from -0.633 to 0.277.

Panels B-D report the estimates for the long-run labour market outcomes, measured in the 1950, 1960, and 1970 censuses. For men, the results show that school closures in general did not affect the long-run outcomes; we find precisely estimated null effects in the DiD specifications for educational attainment, employment, and income. For income, for instance, we can rule out effects larger than -0.9% and +2.2% for men, and -0.4% and +5.2% for women using the DiD estimates and the 95% confidence level.

For women, we do find some evidence that school closures led to a small but statistically significant shift away from retirement (-1.1pp) to employment (+1.4pp) in 1970, which is also reflected in slightly higher incomes. To ensure that this result does not just reflect false rejections due to considering multiple outcomes, we also apply the method proposed by Anderson (2008) to compute the sharpened False Discovery Rate (FDR) q-values. All coefficients become insignificant once we adjust for multiple hypothesis testing–apart from employment, where the coefficient remains significant ($p = 0.048$).[21] Once we condition on being on the labour market, the effect on income becomes even smaller in magnitude and statistically insignificant, indicating that these individuals were not more productive in the labour market (not reported). The additional income from being on the labour market (197 SEK in 1970, which is equal to about 37 U.S. dollars today) is also small in magnitude evaluated at the mean in 1970 (circa 197/8000 = 2.5 percent). We also find no statistically significant effect on educational attainment or occupational score as was also the case for men. All these results hold when varying the observation window and using the difference-in-discontinuities models instead, see Appendix Tables A.3-A.4.

In a supplementary analysis, we also run the analysis comparing children in grade 3 (born in 1909) with those in grade 4 (born in 1908) at the time of the school closure. We do this as choices about their further educational pathways were typically made in grade 3. Thus, children in grade 4 had taken their decision in the year before the pandemic, and

[21] We also apply the method to men's outcomes, and confirm that none of the coefficients are statistically significantly different from zero.

children about to take the decision were directly affected by the pandemic. Interruptions in grade 3 could therefore be expected to have longer-term results compared to interruptions in grade 4. However, we also do not find any effects long-run effects, see Appendix Table A.5. This is consistent with the visual evidence presented in Figure 8 where we also not observe any large discontinuities by birth cohorts.[22] We also checked and found no substantially different patterns by school type as a proxy for school quality or by length of closure.[23]

6 Conclusion

Our paper contributes new insights into the effects of school closures during pandemics, offering a comprehensive short- and long-run analysis of mortality and socio-economic outcomes following the 1918 influenza pandemic in Sweden. Our study stands out by encompassing a nationwide perspective and leveraging individual death certificates, transcribed using machine learning and hand-writing recognition technology. To the best of our knowledge, this is the first investigation to employ weekly and daily panel models, covering a five-year period from 1915 to 1920, to examine school closures during the 1918 Pandemic at the finest geographical level of aggregation. Moreover, we integrate our newly collected school-closure data with unique Swedish individual-level data, enabling a comprehensive examination of the long-term effects on affected school children.

Our analysis yields three key insights. First, we find that schools were closed in response to the worsening local epidemic, highlighting a reactive approach to mitigating the spread of the disease. Second, our evidence indicates that timely implementation was successful in flattening the epidemic curve. We observe that fast (vs slow) implementation of closures reduced peak-level mortality rates, while it did not significantly impact overall mortality rates

[22] We find similar results for children in their last year of primary versus those who had just left primary school. Findings available upon request.

[23] Results are available on request.

evaluated 20 weeks after closure. This finding indicates that the effects of fast closures are concentrated in the short term. Moreover, our examination by age groups reveals a distinct pattern: older age groups initially experienced a preservation of life due to fast closures but subsequently witnessed increased mortality rates, whereas individuals in younger age groups (14-49) experienced life-saving benefits that persisted in the long run. We conjecture that this difference is due to the age-specific mortality of the initial virus and its mutations into less harmful versions as the pandemic progressed. Third, the short-term school closures did not lead to substantial long-term labour market costs for the children affected. Importantly, our precise estimates rule out the possibility of relatively small-sized effects of the closures. For instance, with 95 percent certainty, we can rule out effects on longevity larger than -1% and +0.3% for men, and -.9% and +0.4% for women, in our most precise DiD specifications. The findings are robust to alternative estimation windows and estimation strategies and do not suffer from survivorship bias as we show.

Our study offers valuable lessons derived from these findings. Firstly, despite the relatively short duration of closures, our evidence suggests that faster implementation successfully flattened the epidemic curve, illustrating the potential effectiveness of school closures in mitigating the spread of communicable diseases. Moreover, we discover that when the mortality-age profile of an epidemic disproportionately affects prime-age adults, fast closures can ultimately save lives in the long run rather than merely postponing death for a shorter period. Furthermore, our findings indicate that there were no substantial long-term labour market costs for school children who were sent home during the pandemic. It is important to note that the short average closure duration limits the generalisability of these insights to individual school-absenteeism, as our quasi-experiment involved the complete closure of schools, thus not capturing potential peer-effects. Consequently, our study implies that short-lived, fast school closures can effectively control an epidemic without imposing significant relative long-term costs on the affected primary aged-school children.

A natural question to ask is what we can learn from this experience for more recent events as the context of 1918 differs substantially from today–the population of students likely differs, schooling methods (such as remote teaching) have changed, closures were shorter in 1918 than in most countries during the Covid pandemic,[24] and the mortality rate among young students was much higher during the 1918 pandemic than during more recent pandemics (Ager et al., 2022). Due to these contextual differences, we would argue to be cautious inferring from these results that closing schools nowadays, as happened during the Covid pandemic, will not harm children in the long-run. Despite the possibilities of substituting in-class with remote teaching, there is mounting evidence indicating that school closures during Covid harmed children's mental health, reduced their learning, and increased socio-economic inequalities (e.g. see Betthäuser et al., 2023; Viner et al., 2022). This difference makes sense as school closures during Covid were much longer compared to the school closures during the time period that we examine.

At the same time, our findings for mortality still appear quite timely with contemporary relevance, as we show that curbing an epidemic spread at an early stage can help to reduce the epidemic intensity. Importantly, the short-term school closures did not lead to substantial long-term labour market costs for the children affected. This indicates that with careful implementation, such interventions can mitigate the spread of disease without causing negative socio-economic consequences in the long-run. These results highlight once more the timeless, fundamental trade-off that arises during pandemics: weighing immediate benefits of reduced disease transmission against potential long-term costs. Our study underscores the point that policy makers need to consider this trade-off and to carefully think about a) alternative means to achieve the benefits, b) alternative ways of reducing the costs, and c) the weighting of short-term benefits against the potential long-term costs.

[24]The longest school closure that we observe in our data, at 13 weeks, corresponds to percentile 43 in the international distribution of school closures during the Covid pandemic (UNESCO, 2022).

References

Adda, Jérôme, "Economic activity and the spread of viral diseases: Evidence from high frequency data," *The Quarterly Journal of Economics*, 2016, *131* (2), 891–941.

Ager, Philipp, Katherine Eriksson, Ezra Karger, Peter Nencka, and Melissa A Thomasson, "School closures during the 1918 flu pandemic," *Review of Economics and Statistics*, 2022, pp. 1–28.

Almond, Douglas, "Is the 1918 influenza pandemic over? Long-term effects of in utero influenza exposure in the post-1940 US population," *Journal of Political Economy*, 2006, *114* (4), 672–712.

Alsan, Marcella and Claudia Goldin, "Watersheds in child mortality: The role of effective water and sewerage infrastructure, 1880–1920," *Journal of Political Economy*, 2019, *127* (2), 586–638.

Åman, Margareta, "Spanska sjukan: den svenska epidemin 1918-1920 och dess internationella bakgrund." PhD dissertation, Acta Universitatis Upsaliensis 1990.

Anderson, Michael L, "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American statistical Association*, 2008, *103* (484), 1481–1495.

Aucejo, Esteban M and Teresa Foy Romano, "Assessing the effect of school days and absences on test score performance," *Economics of Education Review*, 2016, *55*, 70–87.

Barro, Robert J, "Non-pharmaceutical interventions and mortality in US cities during the great influenza pandemic, 1918–1919," *Research in Economics*, 2022, *76* (2), 93–106.

_ , **Jose F Ursua, and Joanna Weng**, "The Coronavirus and the Great Influenza Pandemic: Lessons from the Spanish Flu for the Coronavirus Potential Effects on Mortality and Economic Activity," Technical Report, National Bureau of Economic Research 2020.

Beach, Brian, Karen Clay, and Martin Saavedra, "The 1918 influenza pandemic and its lessons for COVID-19," *Journal of Economic Literature*, 2022, *60* (1), 41–84.

_ , **Ryan Brown, Joseph Ferrie, Martin Saavedra, and Duncan Thomas**, "Reevaluating the long-term impact of in utero exposure to the 1918 influenza pandemic," *Journal of Political Economy*, 2022, *130* (7), 1963–1990.

Berkes, Enrico, Olivier Deschenes, Ruben Gaetani, Jeffrey Lin, and Christopher Severen, "Lockdowns and innovation: Evidence from the 1918 flu pandemic," Technical Report, National Bureau of Economic Research 2020.

Betthäuser, Bastian A, Anders M Bach-Mortensen, and Per Engzell, "A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic," *Nature Human Behaviour*, 2023, *7* (3), 375–385.

Bhalotra, Sonia, Martin Karlsson, Therese Nilsson, and Nina Schwarz, "Infant Health, Cognitive Performance and Earnings: Evidence from Inception of the Welfare State in Sweden," *The Review of Economics and Statistics*, 03 2021, pp. 1–46.

Boberg-Fazlic, Nina, Maryna Ivets, Martin Karlsson, and Therese Nilsson, "Disease and fertility: Evidence from the 1918–19 influenza pandemic in Sweden," *Economics & Human Biology*, 2021, p. 101020.

Bootsma, Martin CJ and Neil M Ferguson, "The effect of public health measures on the 1918 influenza pandemic in US cities," *Proceedings of the National Academy of Sciences*, 2007, *104* (18), 7588–7593.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting event study designs: Robust and efficient estimation," *arXiv preprint arXiv:2108.12419*, 2021.

Brauner, Jan M, Sören Mindermann, Mrinank Sharma, David Johnston, John Salvatier, Tomáš Gavenčiak, Anna B Stephenson, Gavin Leech, George Altman, Vladimir Mikulik et al., "Inferring the effectiveness of government interventions against COVID-19," *Science*, 2021, *371* (6531).

Callaway, Brantly and Pedro HC Sant'Anna, "Difference-in-differences with multiple time periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

_ **and Tong Li**, "Policy evaluation during a pandemic," *Journal of Econometrics*, 2023, *236* (1), 105454.

Card, David, "Estimating the return to schooling: Progress on some persistent econometric problems," *Econometrica*, 2001, *69* (5), 1127–1160.

Cattan, Sarah, Daniel A Kamhöfer, Martin Karlsson, and Therese Nilsson, "The Long-Term Effects of Student Absence: Evidence from Sweden," *The Economic Journal*, 11 2022, *133* (650), 888–903.

Chinazzi, Matteo, Jessica T Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore y Piontti, Kunpeng Mu, Luca Rossi, Kaiyuan Sun et al., "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak," *Science*, 2020, *368* (6489), 395–400.

Cho, Sang-Wook, "Quantifying the impact of nonpharmaceutical interventions during the COVID-19 outbreak: The case of Sweden," *The Econometrics Journal*, 2020, *23* (3), 323–344.

Correia, Sergio, Stephan Luck, and Emil Verner, "Pandemics depress the economy, public health interventions do not: Evidence from the 1918 flu," *The Journal of Economic History*, 2022, *82* (4), 917–957.

Cutler, David and Grant Miller, "The role of public health improvements in health advances: the twentieth-century United States," *Demography*, 2005, *42* (1), 1–22.

_ , Angus Deaton, and Adriana Lleras-Muney, "The determinants of mortality," *Journal of Economic Perspectives*, 2006, *20* (3), 97–120.

Cygan-Rehm, Kamila, "Are there no wage returns to compulsory schooling in Germany? A reassessment," *Journal of Applied Econometrics*, 2022, *37* (1), 218–223.

Dahl, Christian Møller, Casper Worm Hansen, and Peter Sandholt Jensen, "The 1918 epidemic and a V-shaped recession: evidence from historical tax records," *The Scandinavian Journal of Economics*, 2022, *124* (1), 139–163.

Dehning, Jonas, Johannes Zierenberg, F Paul Spitzner, Michael Wibral, Joao Pinheiro Neto, Michael Wilczek, and Viola Priesemann, "Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions," *Science*, 2020, *369* (6500).

Elenev, Vadim, Luis Quintero, Alessandro Rebucci, and Emilia Simeonova, "Staggered adoption of nonpharmaceutical interventions to contain covid-19 across us counties: Direct and spillover effects," *Available at SSRN 3657594*, 2020.

Esteves, Rui, Kris James Mitchener, Peter Nencka, and Melissa A Thomasson, "Do Pandemics Change Healthcare? Evidence from the Great Influenza," Technical Report, National Bureau of Economic Research 2022.

Federation of Swedish Genealogical Societies, *Swedish Death Index 1860-2017*, Farsta, Sweden: Federation of Swedish Genealogical Societies, 2018.

Fischer, Martin, Gawain Heckley, Martin Karlsson, and Therese Nilsson, "Revisiting Sweden's comprehensive school reform: Effects on education and earnings," *Journal of Applied Econometrics*, 2022, *37* (4), 811–819.

_ , Martin Karlsson, and Nikolaos Prodromidis, "The Long-Term Effects of Hospital Deliveries," 2021.

_ , _ , Therese Nilsson, and Nina Schwarz, "The long-term effects of long terms–compulsory schooling reforms in sweden," *Journal of the European Economic Association*, 2020, *18* (6), 2776–2823.

Flaxman, Seth, Swapnil Mishra, Axel Gandy, H Juliette T Unwin, Thomas A Mellan, Helen Coupland, Charles Whittaker, Harrison Zhu, Tresnia Berah, Jeffrey W Eaton et al., "Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe," *Nature*, 2020, *584* (7820), 257–261.

Gardner, John, "Two-stage differences in differences," *arXiv preprint arXiv:2207.05943*, 2022.

Goodman, Joshua, "Flaking out: Student absences and snow days as disruptions of instructional time," Technical Report, National Bureau of Economic Research 2014.

Haber, Noah A, Emma Clarke-Deelder, Avi Feller, Emily R Smith, Joshua A Salomon, Benjamin MacCormack-Gelles, Elizabeth M Stone, Clara Bolster-Foucault, Jamie R Daw, Laura Anne Hatfield et al., "Problems with evidence assessment in COVID-19 health policy impact evaluation: a systematic review of study design and evidence strength," *BMJ open*, 2022, *12* (1), e053820.

Hatchett, Richard J, Carter E Mecher, and Marc Lipsitch, "Public health interventions and epidemic intensity during the 1918 influenza pandemic," *Proceedings of the National Academy of Sciences*, 2007, *104* (18), 7582–7587.

Haug, Nils, Lukas Geyrhofer, Alessandro Londei, Elma Dervic, Amélie Desvars-Larrive, Vittorio Loreto, Beate Pinior, Stefan Thurner, and Peter Klimek, "Ranking the effectiveness of worldwide COVID-19 government interventions," *Nature Human behaviour*, 2020, *4* (12), 1303–1312.

Hsiang, Solomon, Daniel Allen, Sébastien Annan-Phan, Kendon Bell, Ian Bolliger, Trinetta Chong, Hannah Druckenmiller, Luna Yue Huang, Andrew Hultgren, Emma Krasovich et al., "The effect of large-scale anti-contagion policies on the COVID-19 pandemic," *Nature*, 2020, *584* (7820), 262–267.

Imbens, Guido W and Donald B Rubin, *Causal inference in statistics, social, and biomedical sciences*, Cambridge: Cambridge University Press, 2015.

Jensen, Robert, "The (perceived) returns to education and the demand for schooling," *The Quarterly Journal of Economics*, 2010, *125* (2), 515–548.

Karlsson, Martin, Daniel Kühnle, and Nikolaos Prodromidis, "The 1918-1919 Influenza Pandemic in Economic History," *Oxford Research Encyclopedia of Economics and Finance*, 2021.

_ , _ , and _ , "The 1918–1919 Influenza Pandemic in Economic History," in "Oxford Research Encyclopedia of Economics and Finance" 2022.

_ , Therese Nilsson, and Stefan Pichler, "The impact of the 1918 Spanish flu epidemic on economic performance in Sweden: An investigation into the consequences of an extraordinary mortality shock," *Journal of health economics*, 2014, *36*, 1–19.

Klugman, Keith P, Christina Mills Astley, and Marc Lipsitch, "Time from illness onset to death, 1918 influenza and pneumococcal pneumonia," 2009.

Liu, Jing, Monica Lee, and Seth Gershenson, "The short-and long-run impacts of secondary school absences," *Journal of Public Economics*, 2021, *199*, 104441.

Mamelund, Svenn-Erik, "Diffusjon av influensa i Norge under spanskesyken 1918-19," *Norsk Epidemiologi*, 1998, *8* (1).

Markel, Howard, Harvey B Lipman, J Alexander Navarro, Alexandra Sloan, Joseph R Michalsen, Alexandra Minna Stern, and Martin S Cetron, "Nonpharmaceutical interventions implemented by US cities during the 1918-1919 influenza pandemic," *Jama*, 2007, *298* (6), 644–654.

National Board of Health, "Allmän hälso- och sjukvård. 1918," Stockholm 1920.

_ , "Allmän hälso- och sjukvård. 1919," Stockholm 1921.

_ , "Allmän hälso- och sjukvård. 1920," Stockholm 1922.

Olden, Andreas and Jarle Møen, "The triple difference estimator," *The Econometrics Journal*, 03 2022, *25* (3), 531–553.

Rosenzweig, Mark R, "Why are there returns to schooling?," *The American Economic Review*, 1995, *85* (2), 153–158.

Roth, Jonathan, Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe, "What's trending in difference-in-differences? A synthesis of the recent econometrics literature," *Journal of Econometrics*, forthcoming.

Statistics Sweden, "Statistisk årsbok för Sverige 1918.," 1918.

Sun, Liyang and Sarah Abraham, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, 2020.

UNESCO, "UNESCO map on school closures and UIS data," 2022.

Velde, Francois R, "What happened to the US economy during the 1918 influenza pandemic? A view through high-frequency data," *The Journal of Economic History*, 2022, *82* (1), 284–326.

Viner, Russell, Simon Russell, Rosella Saulle, Helen Croker, Claire Stansfield, Jessica Packer, Dasha Nicholls, Anne-Lise Goddings, Chris Bonell, Lee Hudson et al., "School closures during social lockdown and mental health, health behaviors, and well-being among children and adolescents during the first COVID-19 wave: a systematic review," *JAMA pediatrics*, 2022.

Xu, Guo, "Bureaucratic Representation and State Responsiveness during Times of Crisis: The 1918 Pandemic in India," *The Review of Economics and Statistics*, 2021, pp. 1–29.

Table 1: Summary statistics, by school closures

	Close = 0		Close = 1		Differences				
VARIABLES	mean	N	mean	N	Raw	p-value	+controls	p-value	Std. Diff.
Panel A: Socio-economic characteristics									
City (0/1)	0.025	828	0.062	1298	-0.037	0.000	0.000	.	-0.181
Population in 1917	1505.558	828	2970.103	1298	-1.5e+03	0.002	-0.000	.	-0.154
Municipal tax rate 1917	7.995	823	12.782	1285	-4.787	0.506	-5.379	0.460	-0.033
Local tax rate 1917	13.967	819	6.914	1294	7.053	0.432	6.919	0.443	0.032
Factories/1000 capita	0.086	828	0.186	1298	-0.100	0.002	0.010	0.490	-0.140
Factory workers/1000 capita	1.937	828	5.999	1298	-4.062	0.000	-0.538	0.404	-0.178
Labor and capital income/capita	228.768	828	321.289	1298	-92.521	0.000	-60.699	0.000	-0.276
Total production value/capita	18.169	828	63.772	1298	-45.603	0.000	-8.008	0.274	-0.186
Revenue/capita	19.931	828	23.496	1298	-3.565	0.000	-1.168	0.042	-0.206
Expenditure/capita	20.985	828	24.810	1298	-3.825	0.000	-1.183	0.082	-0.192
Assets/capita	55.652	828	69.444	1298	-13.792	0.001	0.118	0.964	-0.147
Debt/capita	22.471	828	33.696	1298	-11.225	0.000	-2.252	0.197	-0.189
Recipients of poverty relief/1000 capita	29.075	828	30.495	1298	-1.420	0.051	-0.321	0.640	-0.087
Poorhouses/1000 capita	13.381	828	12.389	1298	0.992	0.059	0.771	0.142	0.083
Military camp (0/1)	0.002	828	0.021	1298	-0.018	0.000	-0.005	0.216	-0.172
Train station (0/1)	0.168	828	0.235	1298	-0.067	0.000	-0.060	0.001	-0.168
Panel B: 1917 national election shares									
Conservative party	0.310	807	0.288	1263	0.022	0.019	0.022	0.018	0.106
Liberal party	0.289	807	0.289	1263	0.000	0.980	-0.002	0.845	0.001
Labour party	0.228	807	0.279	1263	-0.050	0.000	-0.046	0.000	-0.265
Agricultural party	0.128	807	0.089	1263	0.038	0.000	0.035	0.000	0.236
Communist party	0.036	807	0.047	1263	-0.011	0.010	-0.010	0.020	-0.119
Panel C: Health care and mortality									
Hospital (0/1)	0.017	828	0.037	1298	-0.020	0.007	-0.000	0.956	-0.124
Infant mortality rate 1916	68.985	822	66.081	1287	2.904	0.417	2.847	0.429	0.035
Nr. of physicians per 1000	0.122	828	0.145	1295	-0.023	0.007	-0.006	0.449	-0.121
Nr. of nurses per 1000	0.134	828	0.157	1295	-0.023	0.000	-0.016	0.003	-0.176
All-cause mortality rate, 1917	14.902	828	14.283	1298	0.619	0.031	0.593	0.040	0.091

NOTE.— This table reports summary statistics at the parish level. "Close = 0" indicates non-closing parishes (control) and "Close = 1" indicates school closure during the epidemic (treatment). "N" is the number of parishes for which the information is available. All variables are measured prior to the epidemic in 1916 or 1917 (see Section 3 for more details). Column 'raw' reports the raw mean difference, '+controls' reports the adjusted group mean difference controlling for city and population size in 1917. The final column reports the standardised difference.

Table 2: Summary statistics, by response time of school closures

	Slow closures		Fast closures		Differences				
VARIABLES	mean	N	mean	N	Raw	p-value	+controls	p-value	Std. Diff.
Panel A: Socio-economic characteristics									
City (0/1)	0.119	546	0.005	548	0.114	0.000	0.000	.	0.483
Population in 1917	3763.813	546	1643.440	548	2120.373	0.000	0.000	.	0.281
Municipal tax rate 1917	20.791	539	6.965	546	13.826	0.309	16.343	0.246	0.062
Local tax rate 1917	11.418	545	3.566	547	7.852	0.293	8.691	0.259	0.064
Factories/1000 capita	0.361	546	0.010	548	0.351	0.000	0.020	0.344	0.453
Factory workers/1000 capita	11.713	546	0.377	548	11.336	0.000	0.147	0.894	0.392
Labor and capital income/capita	376.694	546	261.431	548	115.263	0.000	29.984	0.096	0.341
Total production value/capita	129.111	546	3.627	548	125.484	0.000	0.662	0.958	0.388
Revenue/capita	27.643	546	19.803	548	7.839	0.000	0.882	0.339	0.384
Expenditure/capita	28.919	546	20.501	548	8.418	0.000	0.847	0.391	0.382
Assets/capita	89.462	546	49.378	548	40.083	0.000	-0.703	0.833	0.408
Debt/capita	47.136	546	21.938	548	25.198	0.000	-2.604	0.257	0.375
Recipients of poverty relief/1000 capita	33.349	546	28.073	548	5.276	0.000	2.354	0.013	0.323
Poorhouses/1000 capita	11.813	546	13.272	548	-1.459	0.039	-0.896	0.216	-0.125
Military camp (0/1)	0.037	546	0.004	548	0.033	0.000	-0.007	0.324	0.236
Train station (0/1)	0.273	546	0.201	548	0.072	0.005	0.057	0.028	0.170
Panel B: 1917 national election shares									
Conservative party	0.271	533	0.301	531	-0.029	0.020	-0.033	0.011	-0.142
Liberal party	0.288	533	0.287	531	0.001	0.944	0.009	0.402	0.004
Labour party	0.291	533	0.277	531	0.014	0.248	0.001	0.956	0.071
Agricultural party	0.087	533	0.089	531	-0.002	0.818	0.007	0.445	-0.014
Communist party	0.061	533	0.035	531	0.027	0.000	0.026	0.000	0.253
Panel C: Health care and mortality									
Hospital (0/1)	0.066	546	0.007	548	0.059	0.000	0.004	0.687	0.316
Infant mortality rate 1916	65.820	543	64.778	543	1.042	0.781	1.879	0.627	0.017
Nr. of physicians per 1000	0.177	546	0.112	548	0.065	0.000	0.019	0.073	0.335
Nr. of nurses per 1000	0.174	546	0.142	548	0.032	0.000	0.024	0.002	0.262
All-cause mortality rate, 1917	14.061	546	14.609	548	-0.547	0.075	-0.422	0.183	-0.108
Days school closed	16.489	546	19.150	548	-2.661	0.000	-2.779	0.000	-0.218

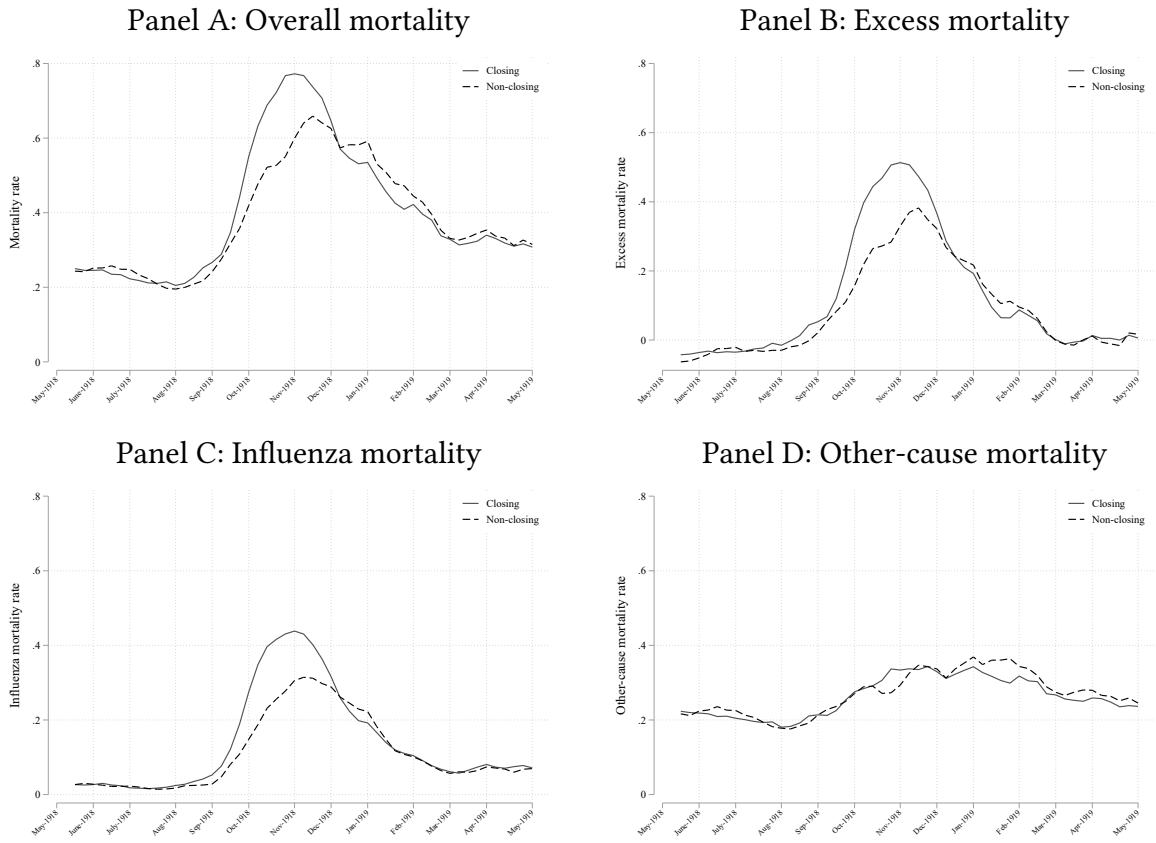
NOTE.— This table reports summary statistics for closing districts by response time ("fast and slow"). All variables except for "days schools closed" are measured prior to the epidemic. See Section 3 for more details. Column 'raw' reports the raw mean difference, '+controls' reports the adjusted group mean difference controlling for city and population size in 1917. The final column reports the standardised difference.

Table 3: Long-run estimates of school closures

	Men		Women	
	DiD	DiDisc	DiD	DiDisc
Panel A: Mortality outcomes				
Died before age 5	0.000	0.002	-0.001	-0.003
	(0.002)	(0.005)	(0.002)	(0.005)
Died during pandemic	-0.001	0.000	-0.000	0.002
	(0.001)	(0.002)	(0.001)	(0.002)
Age at death (years)	-0.252	-0.724	-0.178	-0.361
	(0.225)	(0.597)	(0.232)	(0.675)
N	209666	105181	200000	100257
Panel B: 1950 Census				
Responded Census 1950	-0.001	-0.003	-0.004	-0.001
	(0.003)	(0.009)	(0.003)	(0.009)
N	212566	106563	203675	101997
Employment 1950	-0.000	0.001	-0.001	-0.005
	(0.001)	(0.002)	(0.003)	(0.009)
N	183132	92015	177510	89007
HISCAM 1950 (std.)	0.010	-0.003	-0.031	-0.023
	(0.010)	(0.027)	(0.021)	(0.062)
N	171062	85968	40527	19931
Panel C: 1960 Census				
Responded Census 1960	-0.002	0.004	-0.003	-0.002
	(0.004)	(0.009)	(0.003)	(0.010)
N	212566	106563	203675	101997
A levels	0.000	-0.005	-0.001	-0.000
	(0.002)	(0.006)	(0.001)	(0.003)
N	180141	90495	175071	87703
Employment 1960	0.002	-0.002	0.005	0.003
	(0.002)	(0.005)	(0.005)	(0.014)
N	180141	90495	175071	87703
Panel D: 1970 Census				
Responded Census 1970	-0.003	-0.003	-0.005	-0.007
	(0.004)	(0.010)	(0.004)	(0.011)
N	212566	106563	203675	101997
Sec. education 1970	-0.001	-0.008	-0.001	-0.002
	(0.003)	(0.008)	(0.003)	(0.007)
N	126390	84165	125940	83870
Employment 1970	0.001	-0.016	0.014***	0.033**
	(0.004)	(0.012)	(0.005)	(0.013)
N	167829	84264	167422	83922
Retired 1970	-0.001	0.014	-0.011**	-0.023
	(0.004)	(0.010)	(0.005)	(0.016)
N	167829	84264	167422	83922
Total Income 1970 all	168.820	-559.700	197.574*	306.800
	(211.708)	(639.282)	(117.894)	(338.202)
N	167829	84264	167422	83922

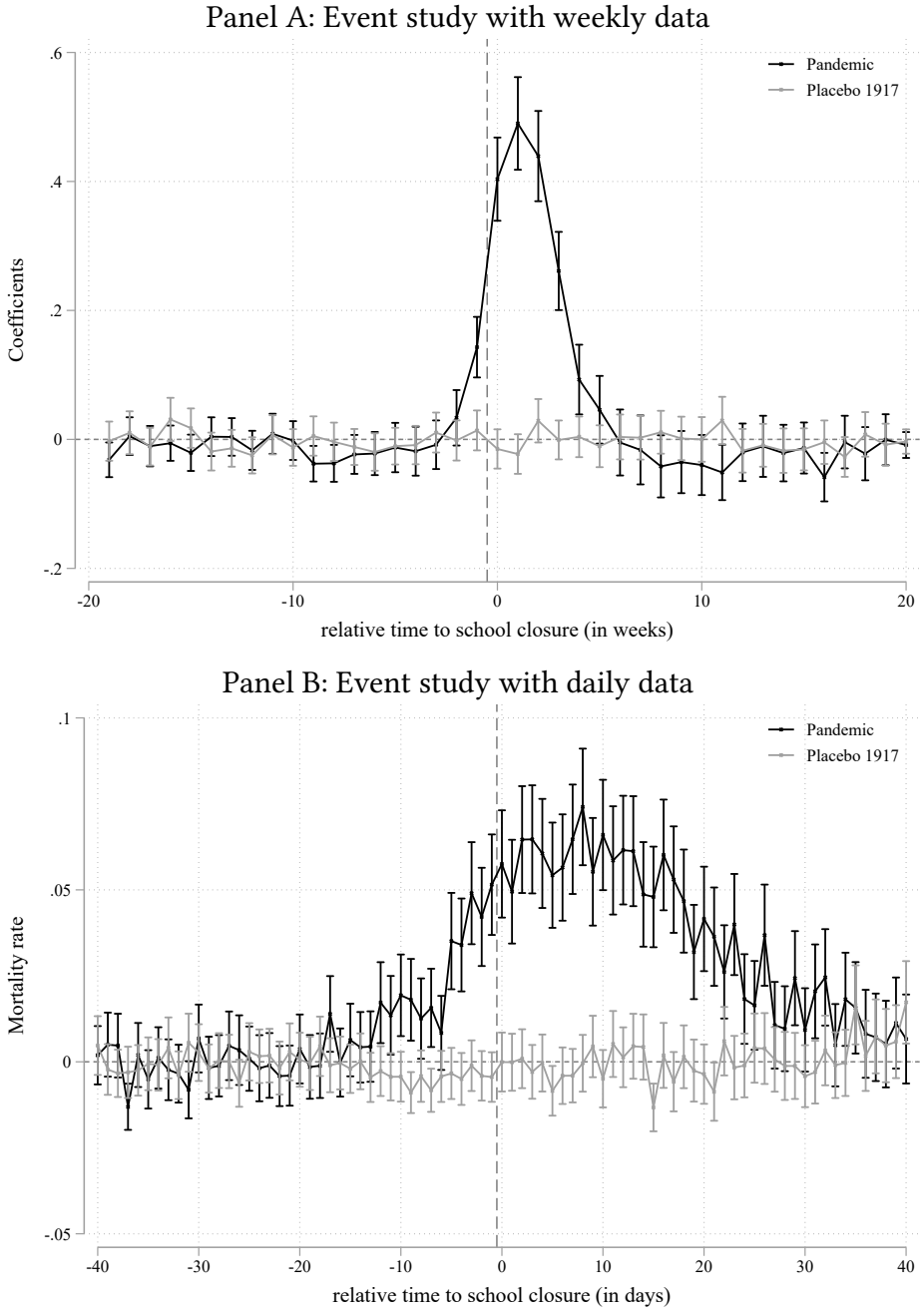
Notes: This table reports the coefficients from the difference-in-differences and difference-in-discontinuities estimations for different outcomes. Each coefficient represents one regression. The DiD regressions uses children born between 1910 and 1913 and fixed effects for parish, birth year and birth month (see Equation 3; standard errors are clustered at the parish level. The DiDisc regressions uses children born between 1911 and 1912 according to Equation 3; standard errors are clustered at the running variable, i.e. day of birth.

Figure 1: Average weekly mortality rates during the epidemic



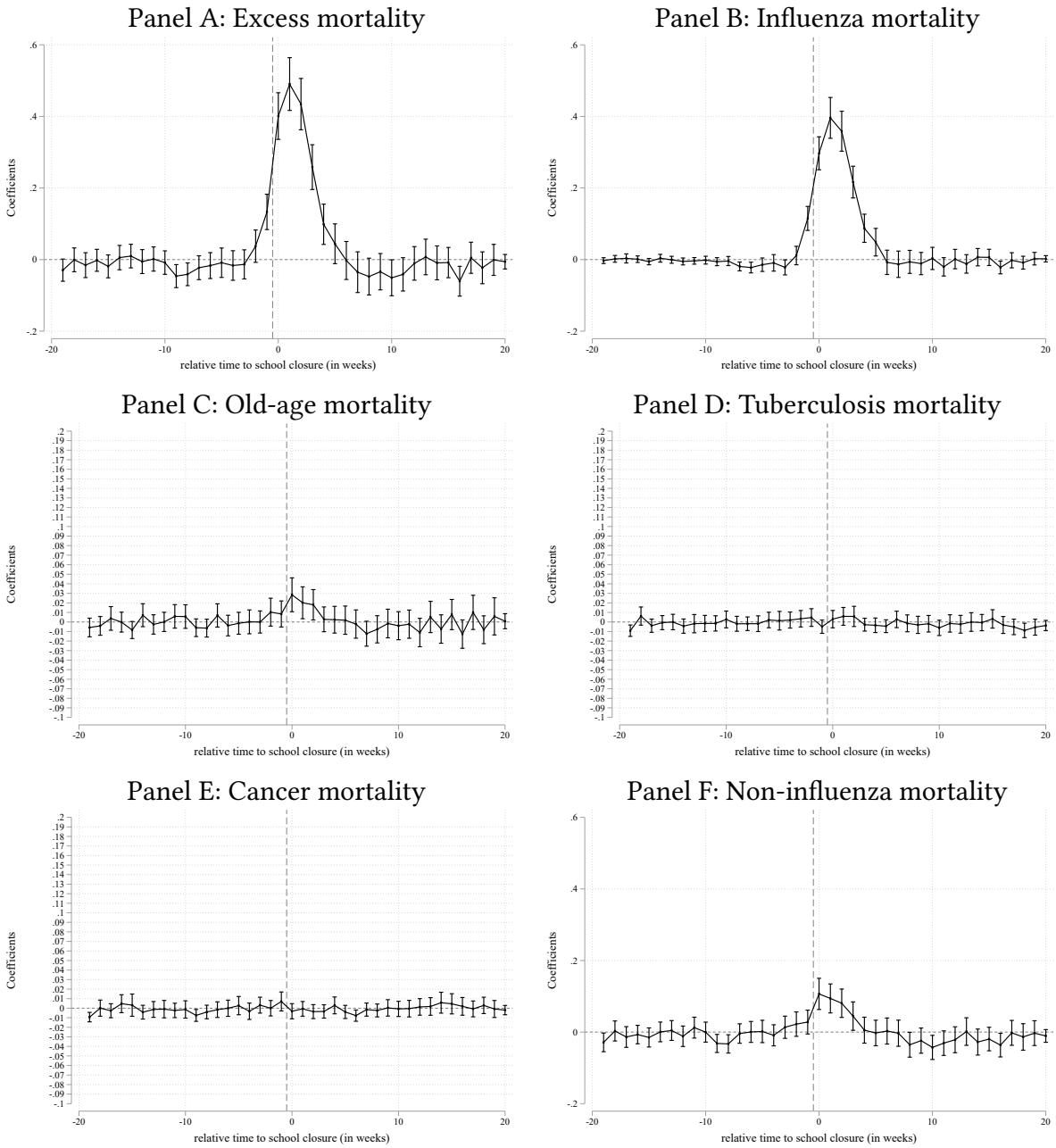
Note: This figure shows weekly average mortality rates (3-week moving averages) for closing- (solid lines) and non-closing (dashed lines) school districts. The all-cause, excess, and influenza mortality rates are reported in Panels A, B, and C, respectively. Panel D plots the mortality rate from all other causes of deaths than influenza (i.e., all death minus influenza deaths).

Figure 2: School closures, baseline findings



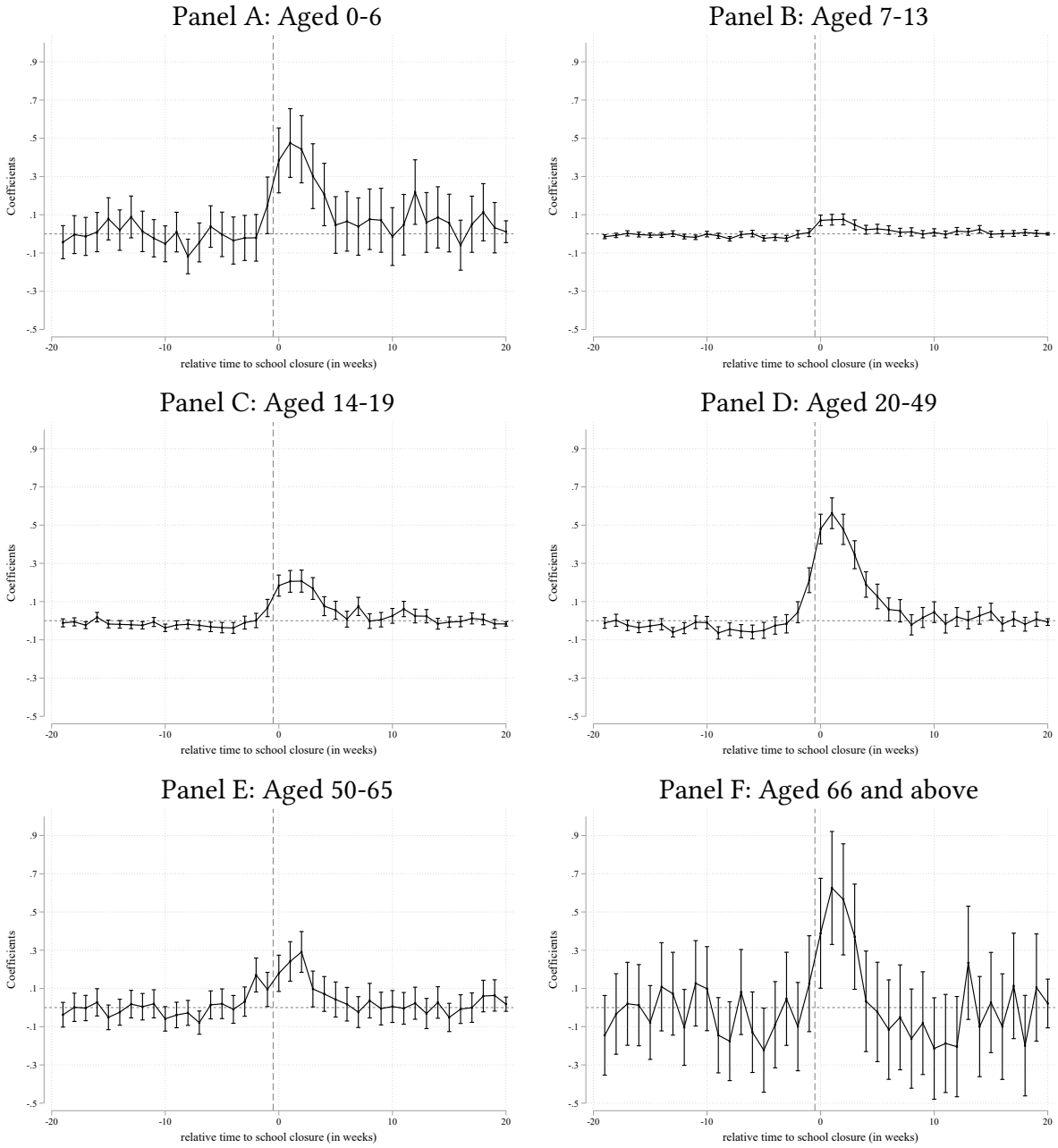
Note: This figure shows the baseline event study coefficients and 95% confidence intervals from estimating Equation 1 using weekly data (Panel A) and daily data (Panel B). The outcome variable is the all-cause mortality rate for the pandemic year (black line) and for the placebo year 1917 (grey). Panel A includes week-date fixed effects, Panel B includes day-fixed effects. In Panel A, the omitted category is 20 weeks before closure (-20). In Panel B, the event-window is 40 days before/after the closure and the omitted category is 40 days before closure (-40).

Figure 3: School closures, alternative mortality outcomes



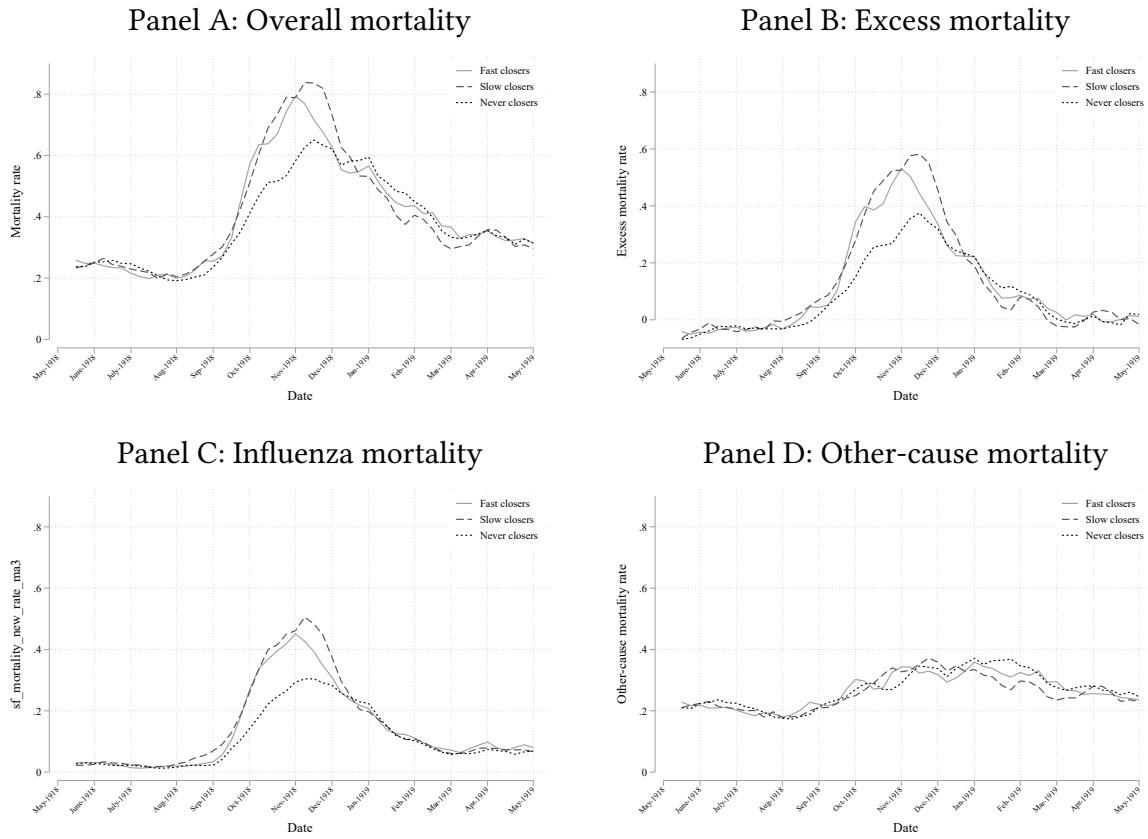
Note: This figure shows event study coefficients and 95% confidence intervals from estimating Equation 1, using different causes of death: The excess mortality rate, which is based on comparing the weekly mortality during epidemic to mortality rates the preceding four years (Panel A); the influenza mortality rate (Panel B); the old-age mortality rate, which are causes of death related to old age, (Panel C); the tuberculosis mortality rate (Panel D); the cancer mortality rate (Panel E); cardiovascular mortality rate (Panel F). The omitted category is 20 weeks before closure (-20).

Figure 4: School closures, age-specific mortality



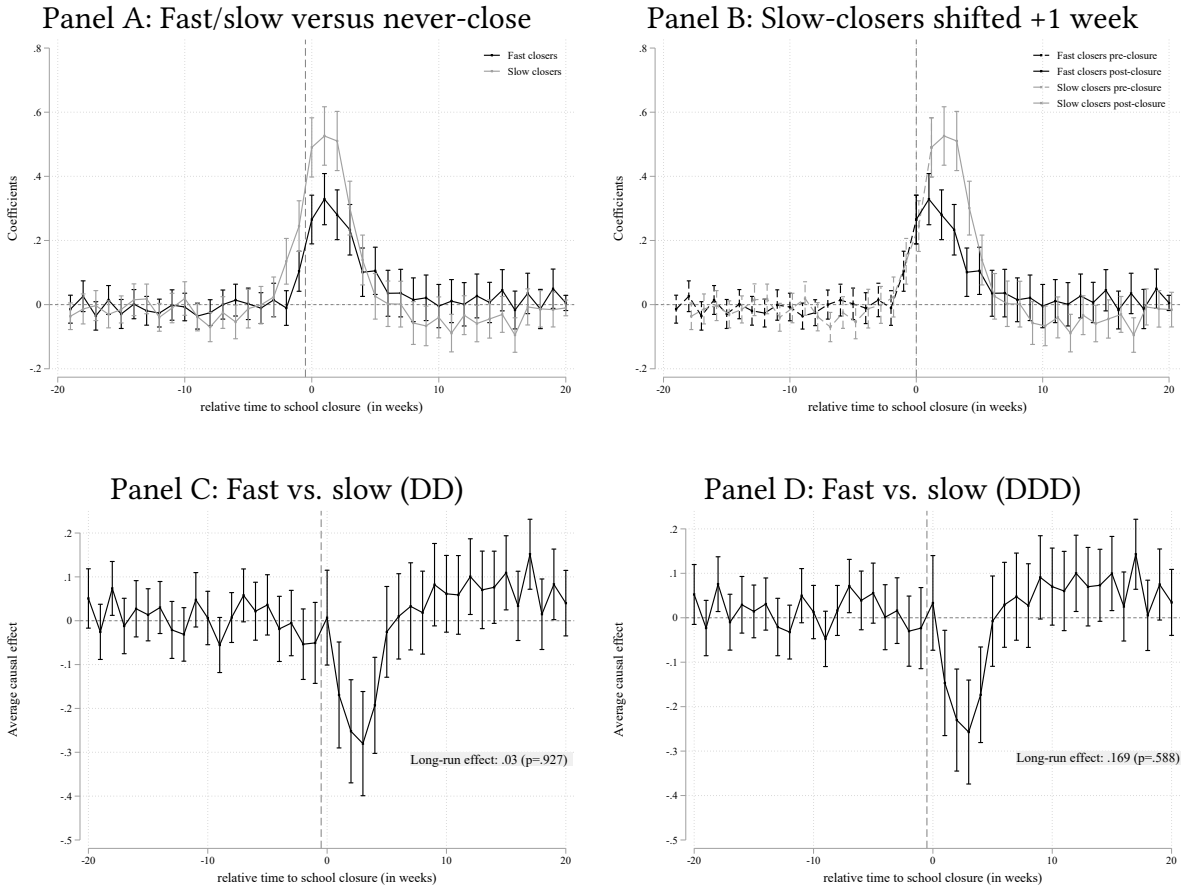
Note: This figure shows event study coefficients and 95% confidence intervals from estimating Equation 1, using different age-specific mortality rates, where the denominator is based on the relevant age groups. Each age-range is indicated in the panel title. The outcomes are winsorised at the 1st and 99th percentile to reduce the influence of outliers on these outcomes. The omitted category is 20 weeks before closure (-20).

Figure 5: Average weekly mortality rates during the epidemic, by response time



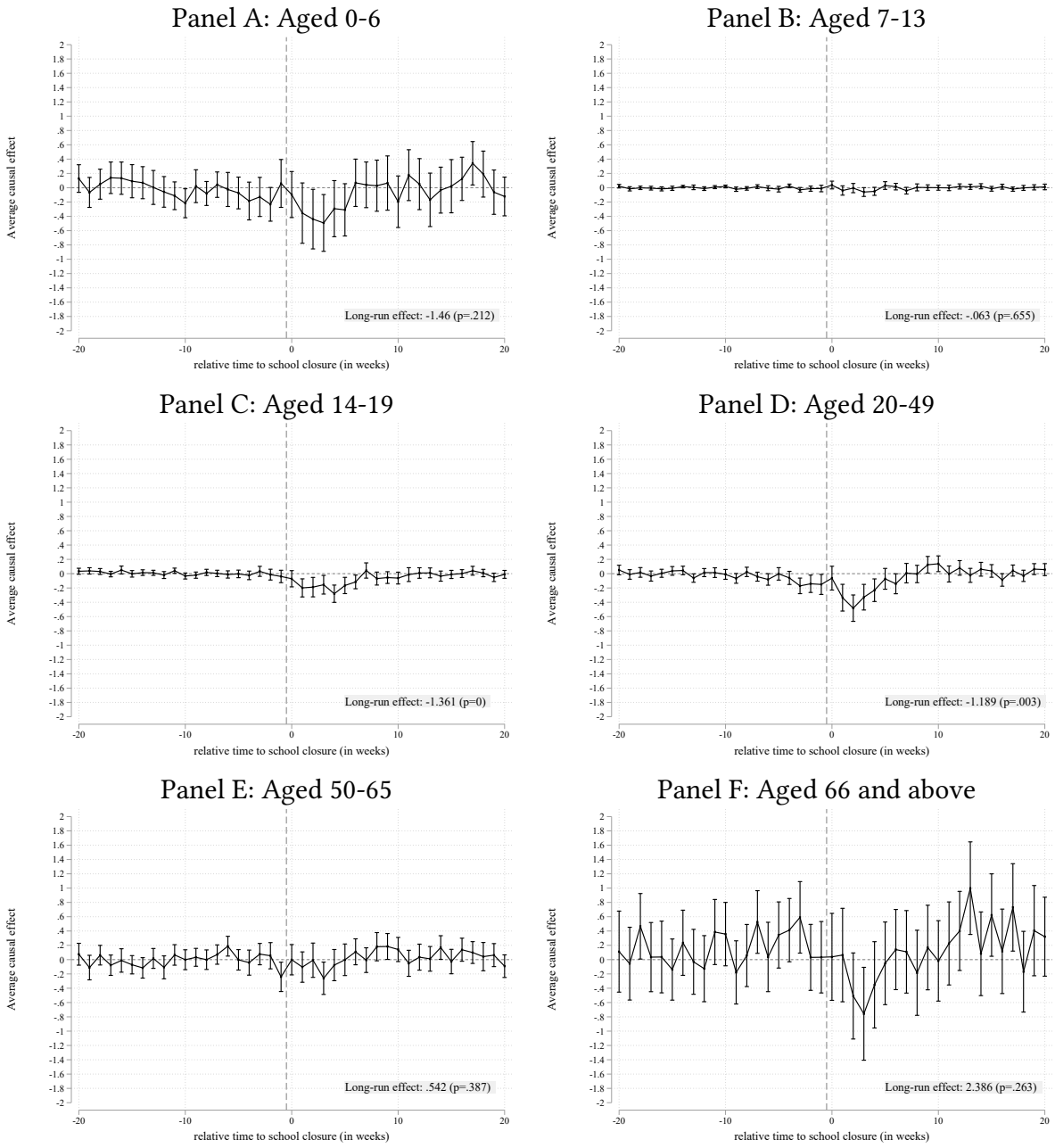
Note: This figure shows weekly average mortality rates (3-week moving averages) for the sample of rural school district by response time and for non-closing districts. The all-cause, excess, and influenza mortality rates are reported in Panels A, B, and C, respectively. Panel D plots the mortality rate from all other causes of deaths than influenza (i.e., all death minus influenza deaths).

Figure 6: Fast closures reduced the intensity of the epidemic



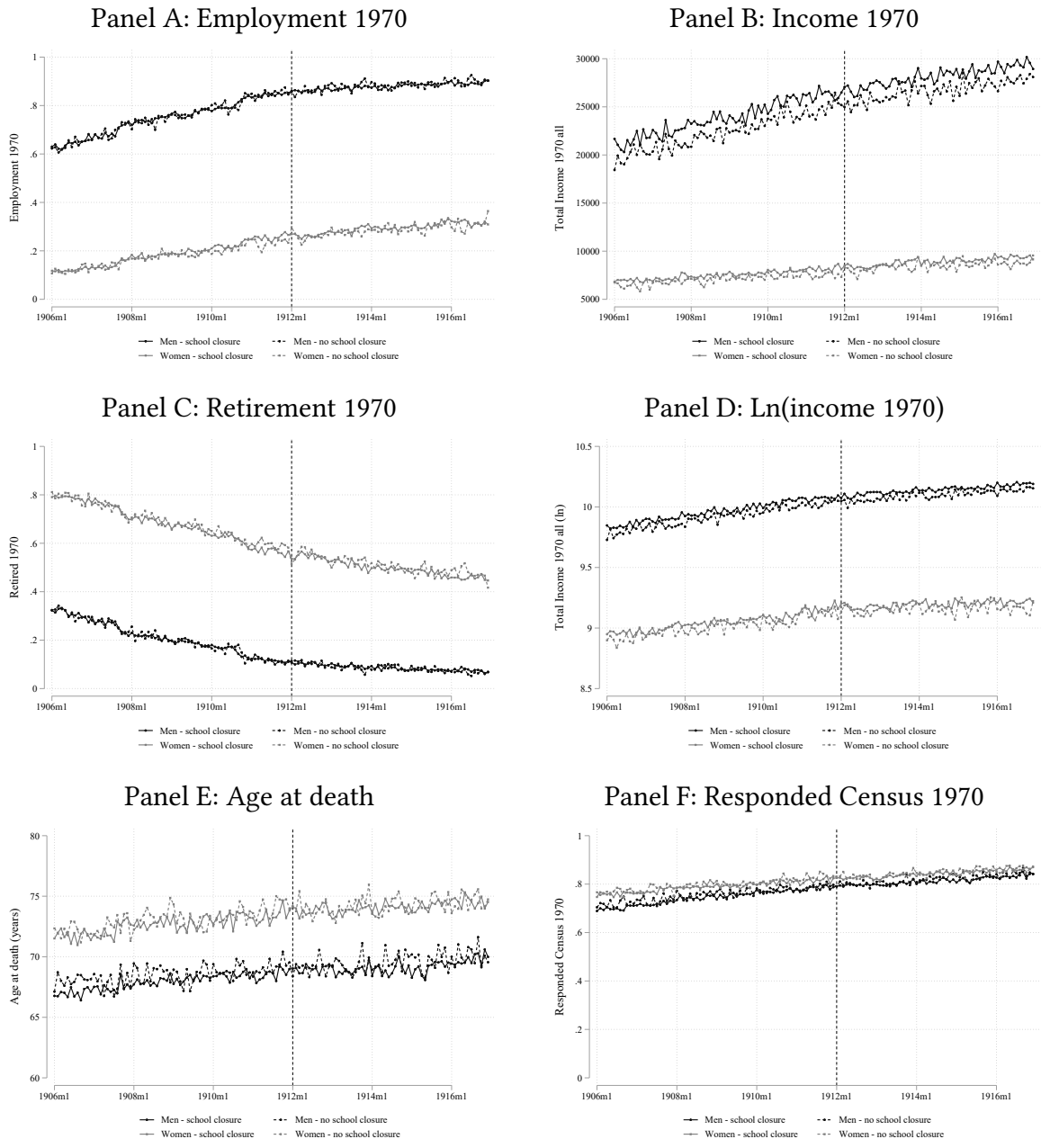
Note: Panel A shows the baseline coefficients and 95% confidence intervals from estimating Equation 1 separately for fast closures vs never closures and slow closures vs never closures. Panel B shows the same estimates, having shifted the closure date of slow closures forwards by one week. Panel C plots the difference between fast vs. and shifted slow closures only. Panel D additionally uses never-closing school districts within a triple-differences design. The outcome variable is the all-cause mortality rate. The long-run effect, reported in Panels C and D, is the sum of all the coefficients after closure. The included districts are from the "rural sample". The omitted category is 20 weeks before closure (-20).

Figure 7: Fast school closures, age-specific mortality



Note: This figure event study estimates as reported in Panel C of Figure 6, but using the age specific mortality rates as outcomes. Each age-range is indicated in the panel title. The long-run effect, reported in each figure, is the sum of all the coefficients after closure. The included districts are from the"rural sample" The omitted category is 20 weeks before closure (-20).

Figure 8: Long-run trends in outcomes



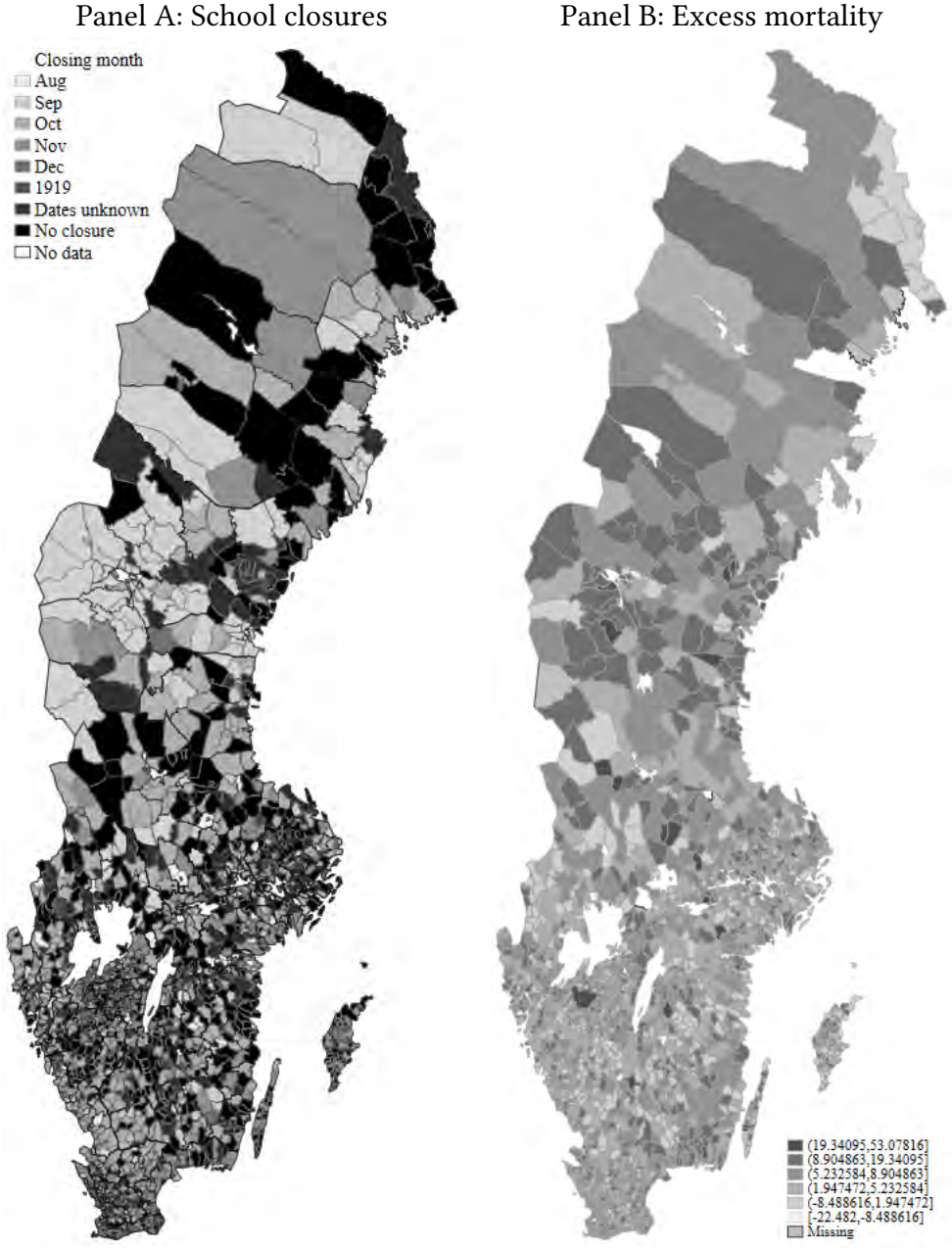
Note: This figure displays the long-run outcomes by sex and by whether they experienced the 1918-epidemic school closures (treatment vs. control). Each observation is the average outcome by birth month. The specific outcome is indicated in each panel title.

Appendices

FOR ONLINE PUBLICATION ONLY

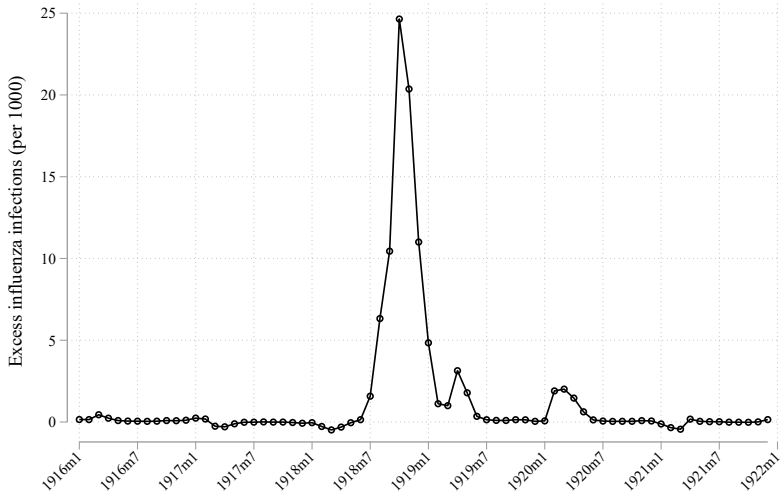
Appendix A Additional tables and figures

Figure A.1: Spatial distribution of school closures and excess mortality



NOTE.— Panel A shows a map over school districts closing schools (color-grouped by closure date) and black-colored districts without school closures. Panel B shows a map of the excess mortality rate, calculated as the difference between the mortality rate during the epidemic period (June 1918 to March 1919) and the average mortality rate during the same months in the previous four years (1914-1917), expressed as the number of deaths per 1,000 people in each parish.

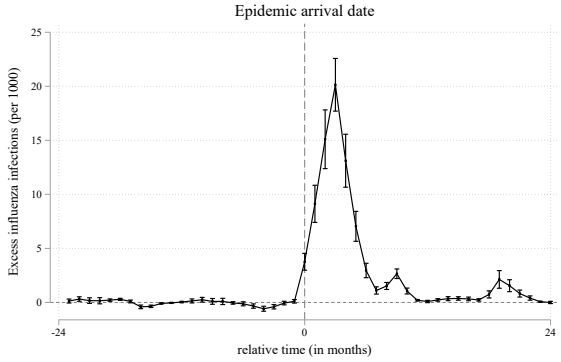
Figure A.2: Monthly excess infections



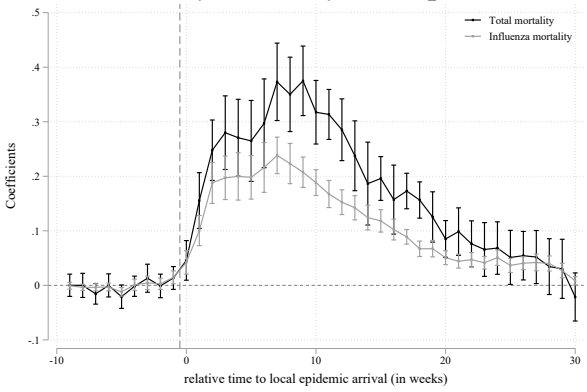
Note.— This figure shows the average monthly excess influenza infections at the health districts level.

Figure A.3: Event study estimates of effect of local arrival

Panel A: Monthly influenza infections, health districts

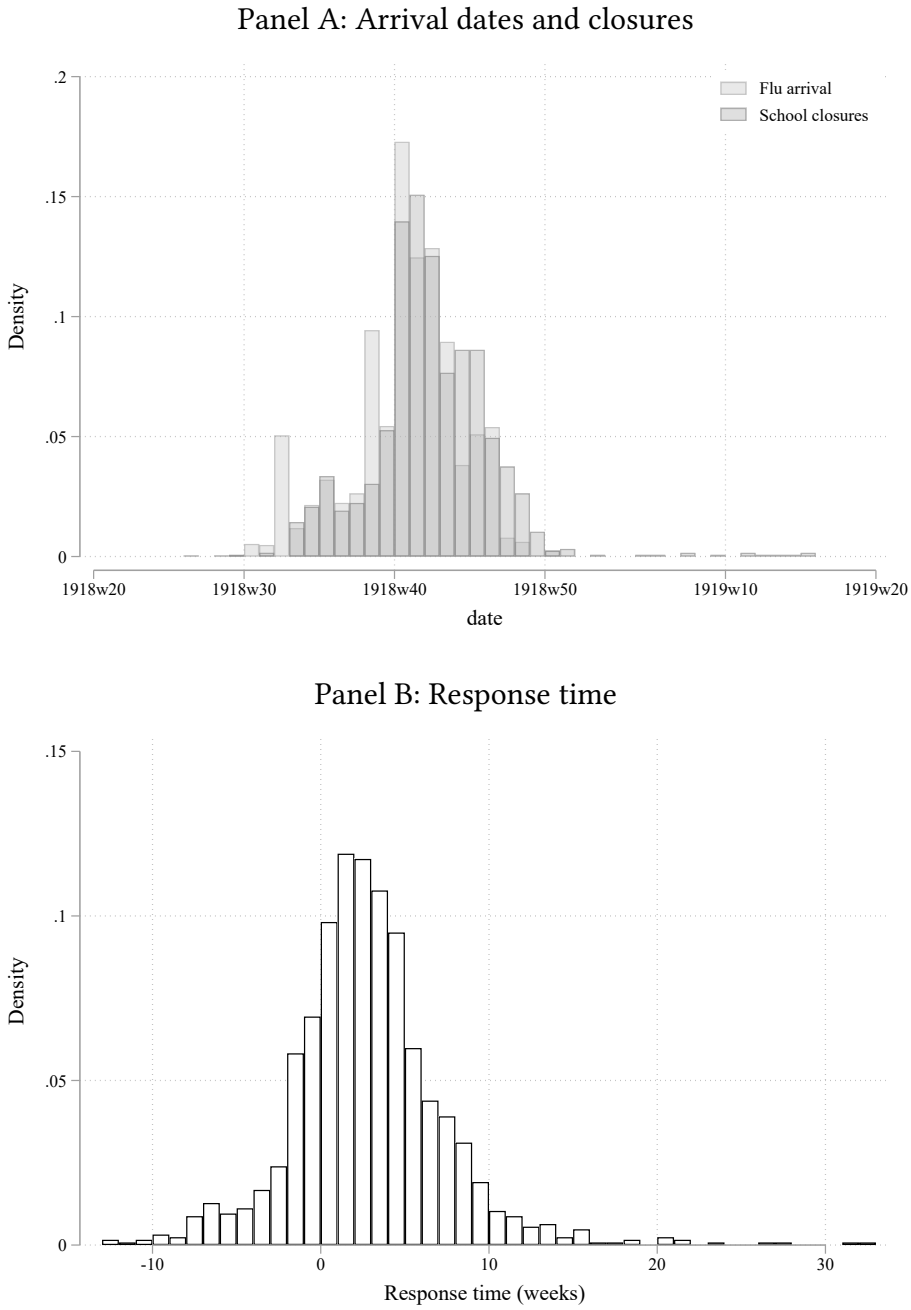


Panel B: Weekly mortality rates, parish-level



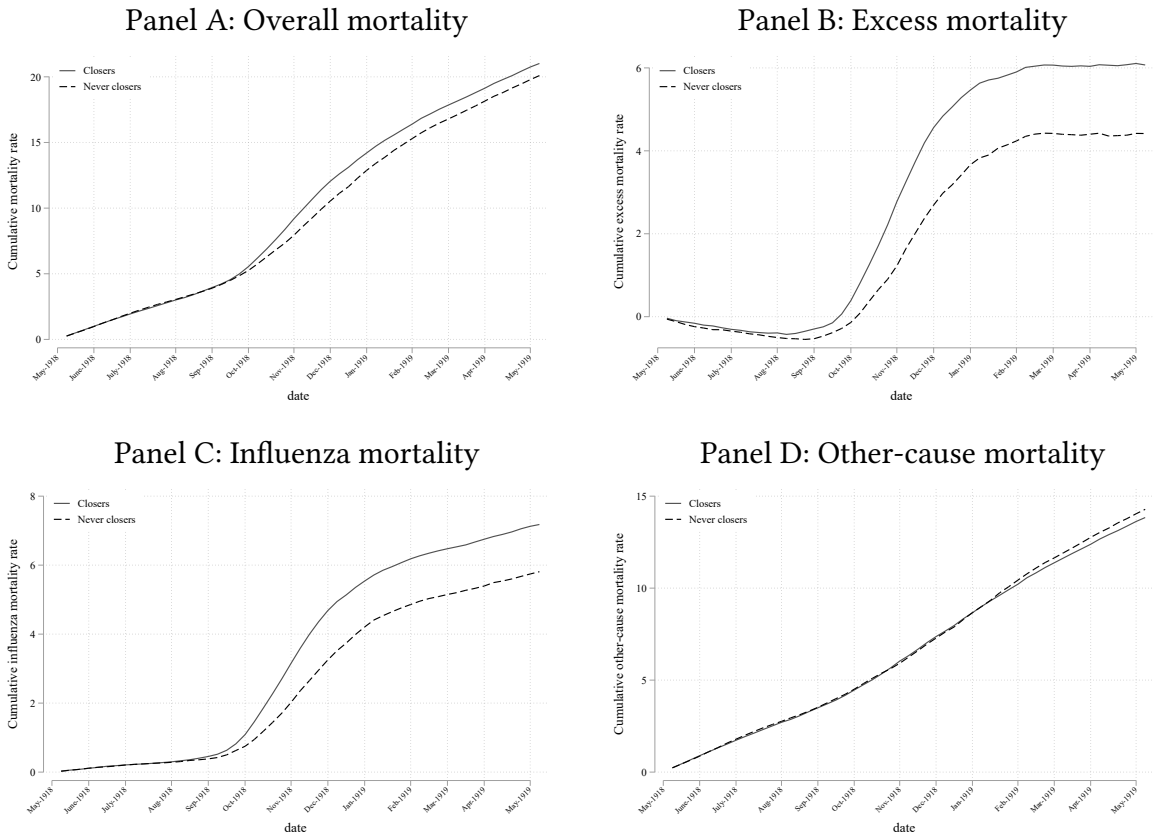
NOTE.— This figure shows TWFE-event study estimates for the effect of the local arrival of the epidemic on the monthly influenza infections rate at the health districts level (Panel A), where we also control district-by-month fixed effects. In Panel B, we report the event study coefficients using the Gardner (2022) estimator to account for dynamic treatment effects on weekly mortality rates (all-causes and influenza) at the parish level (Panel B).

Figure A.4: Distribution of school closures and local epidemic arrival dates



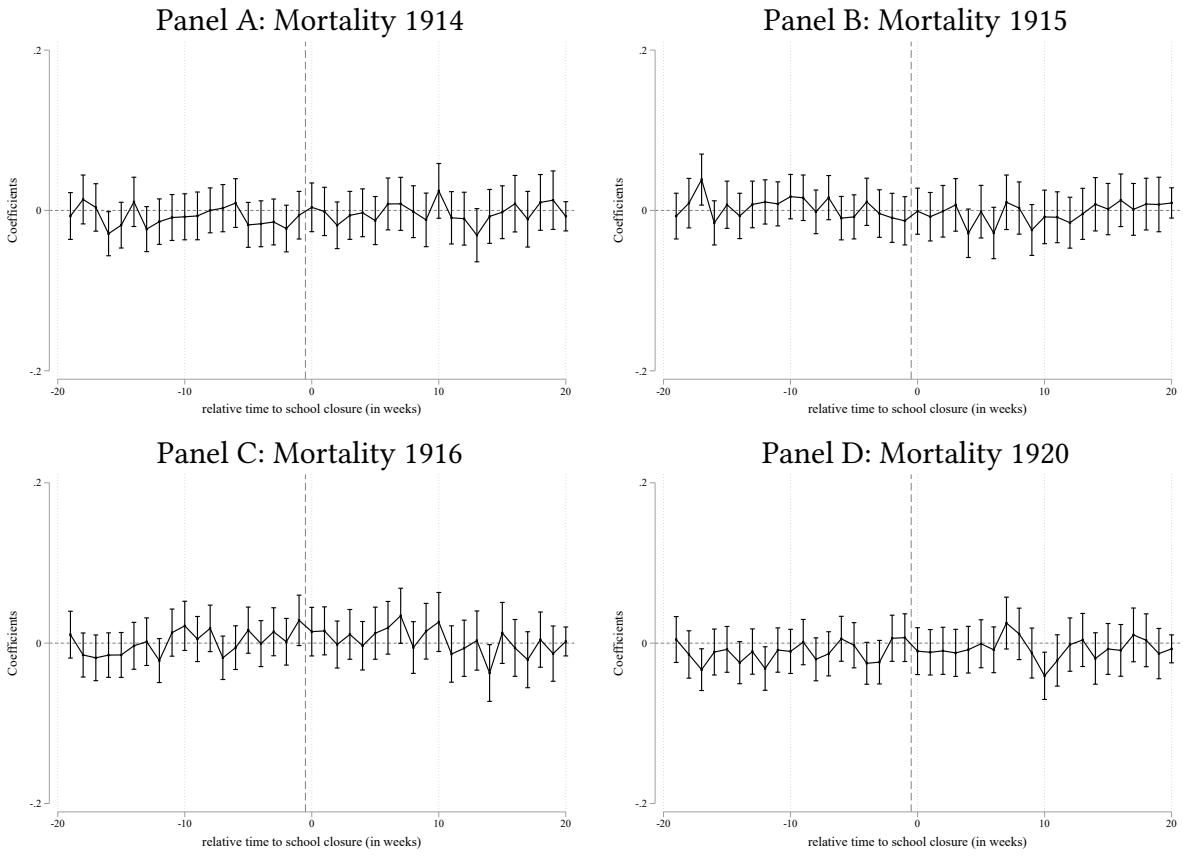
Note: This figure shows the distribution of the estimated local flu arrival dates and the school closure dates (Panel A) and the implied response times (Panel B), calculated as the difference between the week of the school closure minus the week of the epidemic arrival. See Section 3.3 for more details.

Figure A.5: Cumulative weekly mortality rates



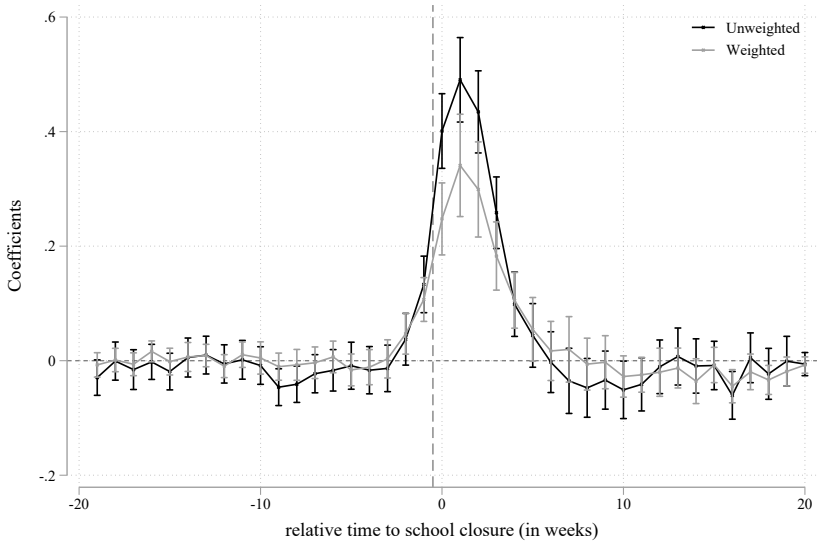
Note: This figure shows weekly cumulative mortality rates for all parishes. Panels A-C depicts the three markers of epidemic mortality: cumulative all-cause mortality rate (black solid), cumulative excess all-cause mortality rate (gray dashed line), and cumulative influenza mortality rate (blue dotted line). Panel D depicts the cumulative all-other causes (than influenza) mortality rate.

Figure A.6: Robustness, placebo years



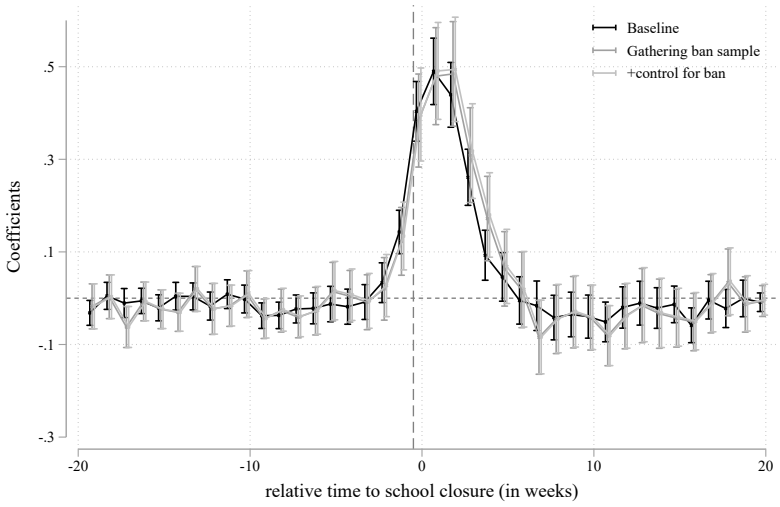
Note: This figure shows the baseline coefficients and 95% confidence intervals from estimating Equation 1 for the weekly all-cause mortality rate from non-pandemic years, where we falsely assume that the school closures happened in the same weeks during those years instead. The omitted category is -20.

Figure A.7: Weighting makes no difference



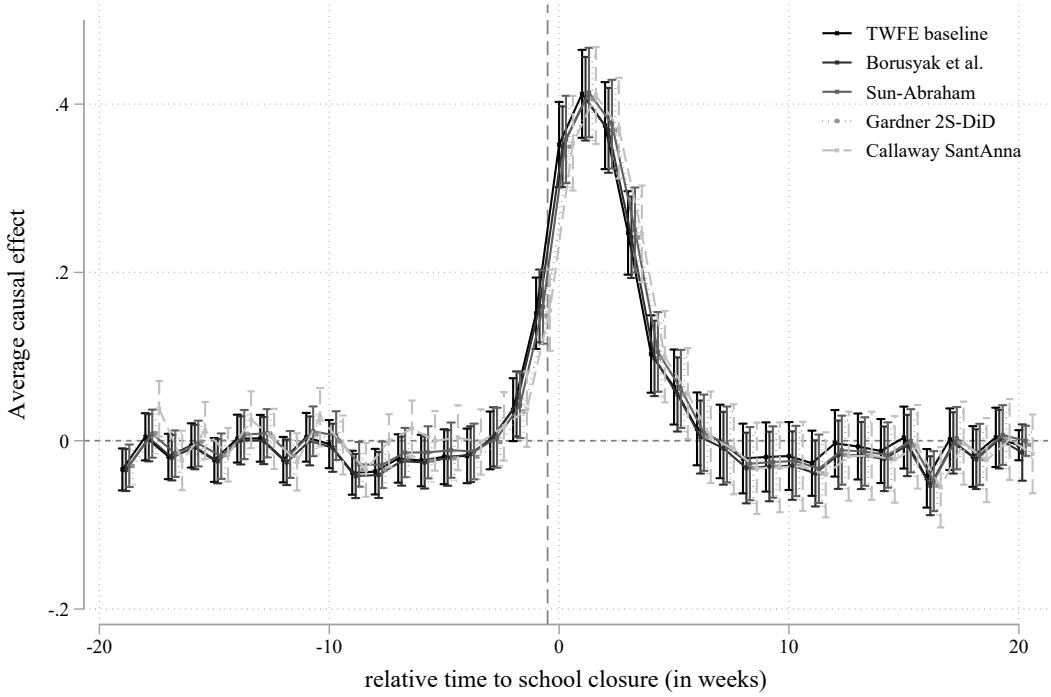
Note.— This figure shows the baseline coefficients and 95% confidence intervals from estimating Equation 1 for the weekly all-cause mortality rate for the unweighted data and using 1917 population weights. The omitted category is -20.

Figure A.8: Controlling for gathering bans makes no difference



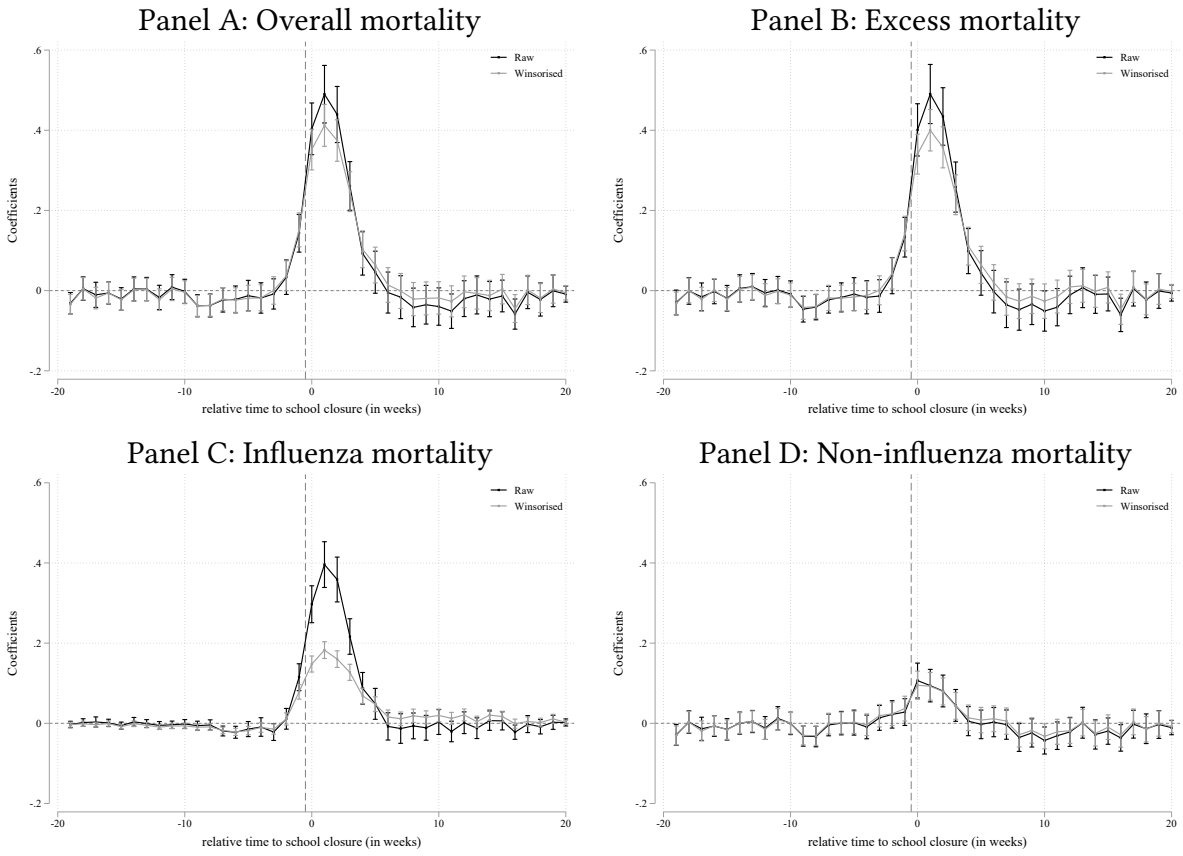
Note.— This figure shows the baseline coefficients and 95% confidence intervals from estimating Equation 1 for the weekly all-cause mortality rate for 1) the baseline sample, 2) parishes where we have complete information about gathering ban dates, 3) and where we control for these dates at the weekly level. The omitted category is -20.

Figure A.9: Robustness dynamic treatment effects



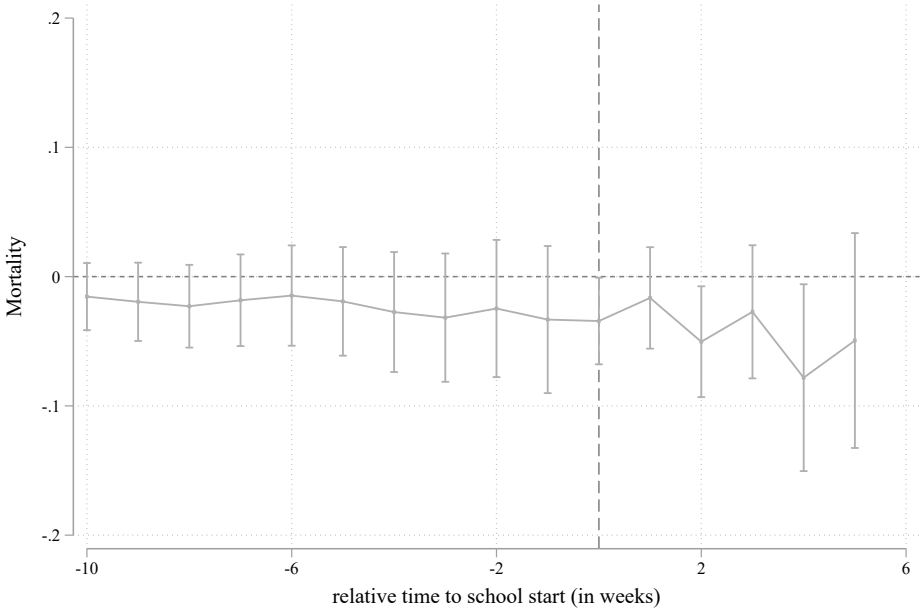
Note: This figure shows that the baseline (weekly) TWFE event study estimates are similar to estimates based on the event study estimates based on Sun and Abraham (2020), Gardner (2022), Callaway and Sant'Anna (2021), and Borusyak et al. (2021) . The outcome variable is the all-cause mortality rate. The omitted category is 20 weeks before closure (-20).

Figure A.10: Robustness to winsorising, alternative mortality outcomes



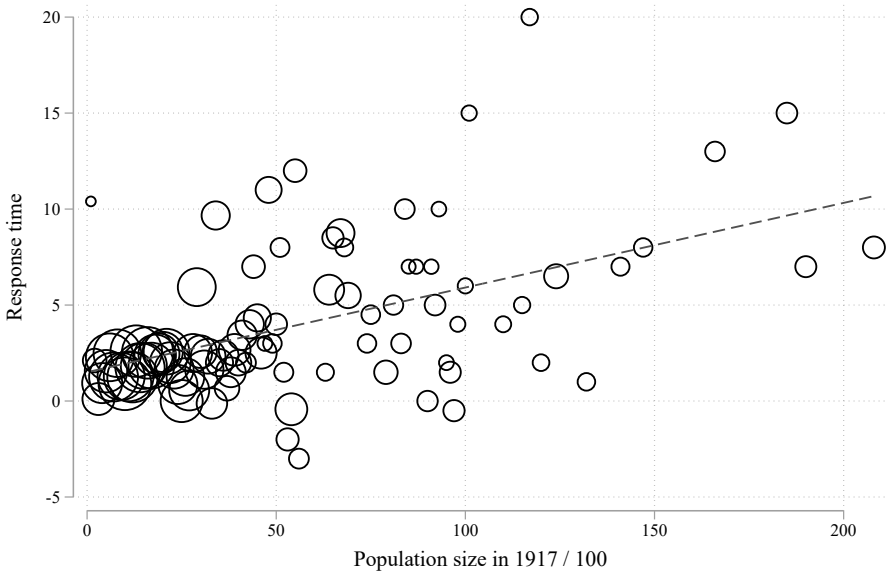
Note: This figure shows event study coefficients and 95% confidence intervals from estimating Equation 1, using the winsorised and non-winsoried variables of different causes of death: The overall mortality rate (Panel A); the excess mortality rate (Panel B); the influenza mortality rate; and all-non-influenza deaths (Panel D). The omitted category is 20 weeks before closure (-20).

Figure A.11: School openings, baseline findings



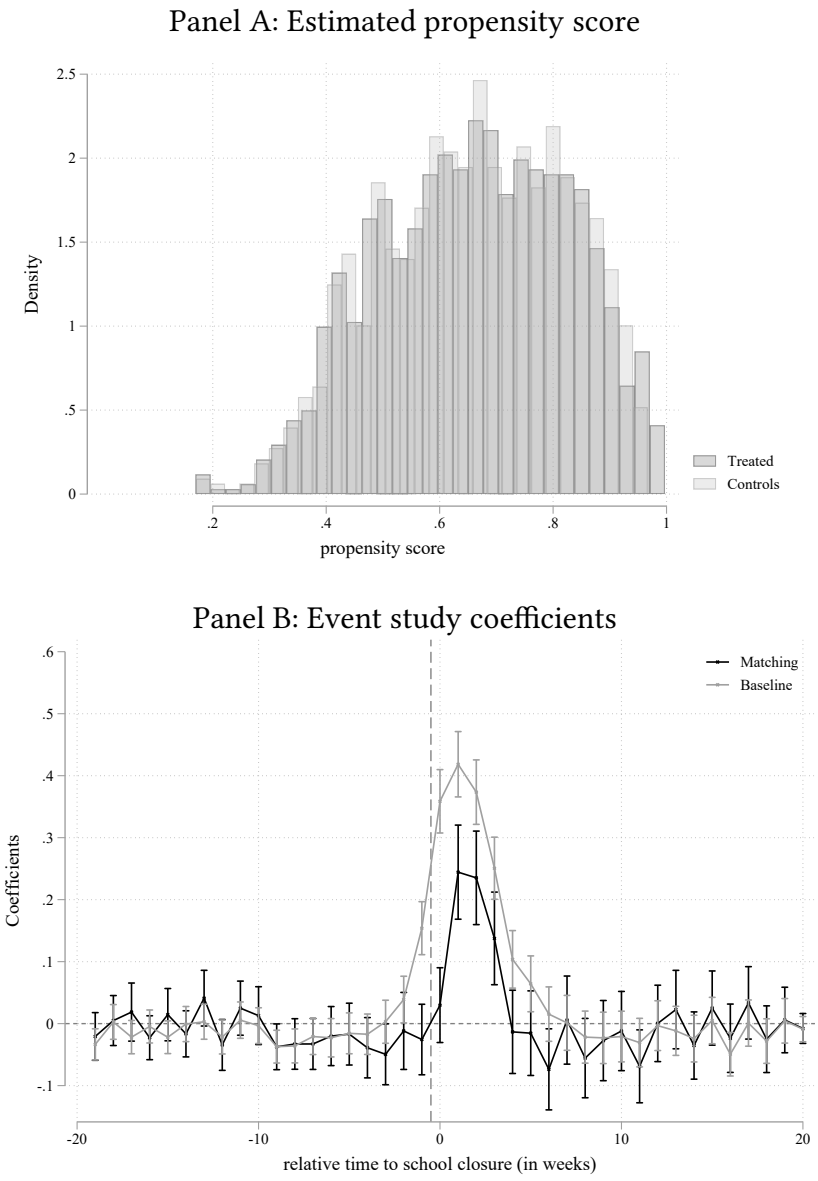
Note: This figure shows the event study coefficients and 95% confidence intervals for the effect of school openings after the summer holidays for districts that opened schools between 1 August 1918 and 16 September 1918. The outcome variable is the all-cause mortality rate. Estimates based on Borusyak et al. (2021) estimator.

Figure A.12: Population size and response speed



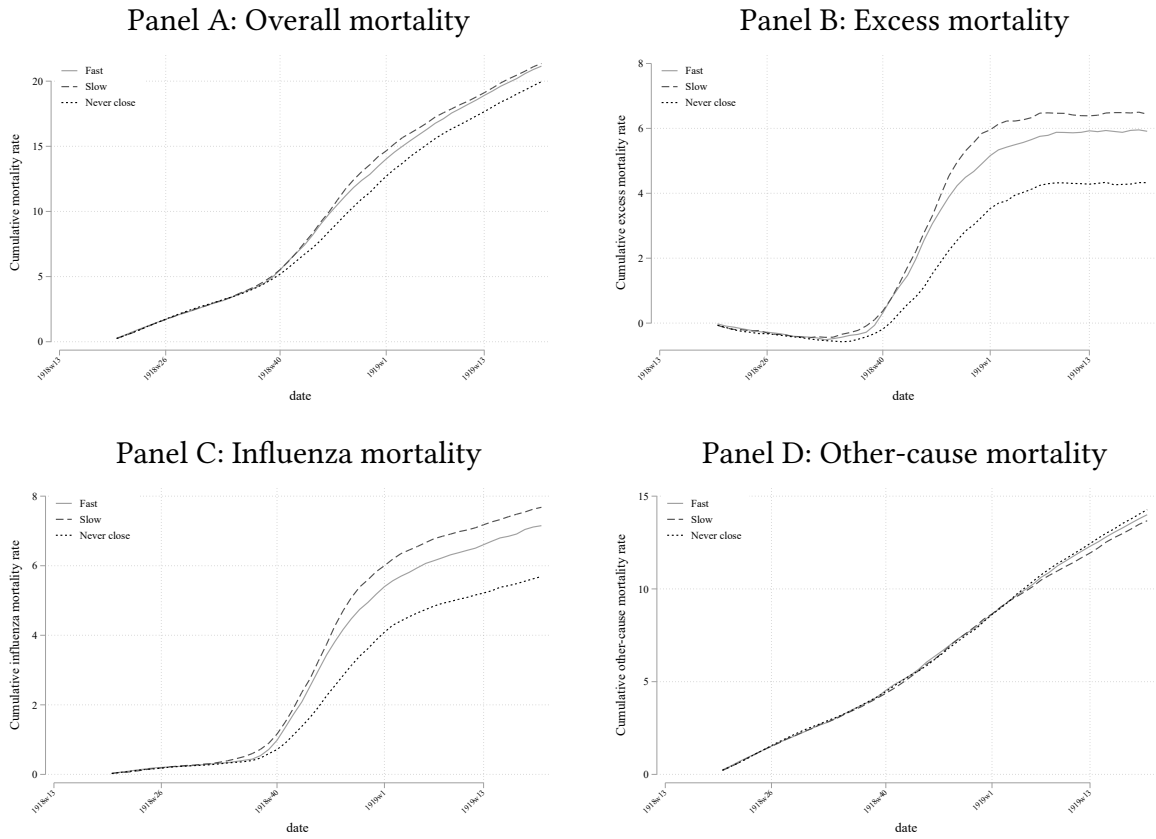
Note.— This figure shows the relationship between population size in rural parishes, binned into groups of 100, and the response time of closing schools measured in weeks. Figures drops cities and focuses on the remaining rural parishes, which corresponds to more than 99% of all parishes. Markers weighted by population size of 1917.

Figure A.13: Matching results



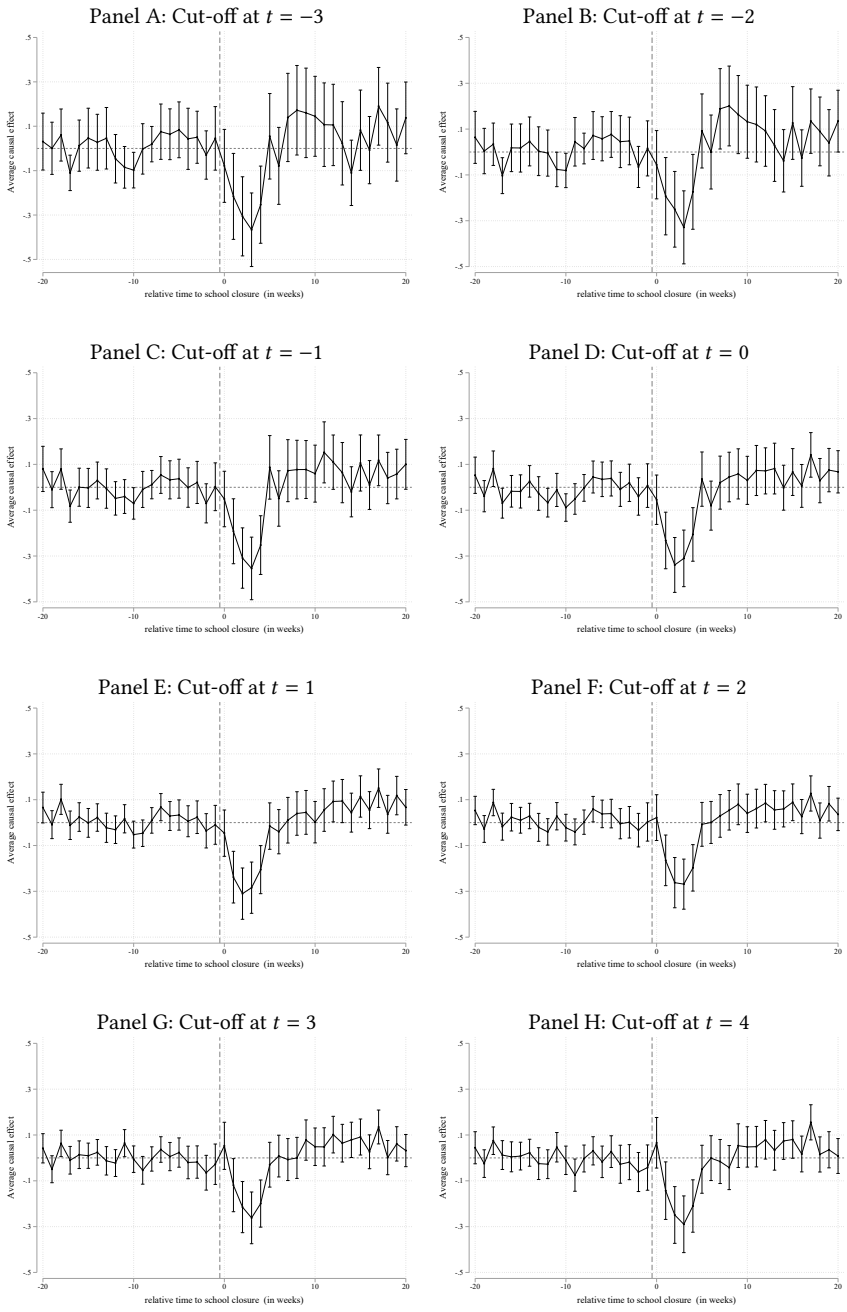
Note.— This figure reports the results from our matching analysis. We combine coarsened exact matching on parish population size with propensity score matching at the parish level; we match closing with non-closing parishes at the point of closure and use socioeconomic characteristics reported in Table 1 as well as pre-closure mortality rates at the daily level as matching variables. We apply a caliper that excludes the worst 50% of matches, and we identify the nearest neighbour. Panel A gives the disrtribution of propensity scores for the matched sample, and Panel B shows the event study coefficients and 95% confidence intervals from estimating Equation 1 for the matched sample, applying inverse probability weighting.

Figure A.14: Cumulative weekly mortality rates, by response time



Note: This figure shows weekly cumulative mortality rates for the rural sample for closing school districts by response time and for non-closing districts. Panels A-C depicts the three markers of epidemic intensity: cumulative all-cause mortality rate (black solid), cumulative excess all-cause mortality rate (gray dashed line), and cumulative influenza mortality rate (blue dotted line). Panel D depicts the cumulative all-other causes (than influenza) mortality rate.

Figure A.15: Sensitivity of fast/slow results for different cut-off values



NOTE.— The figure shows the coefficients and 95% confidence intervals from estimating Equation 1 for fast closures vs slow closures using different cut-off values to define fast and slow closures. The time points in the panels refer to the relative timing of the school closure in weeks relative to the epidemic arrival date. The outcome variable is the all-cause mortality rate. The omitted category is -20.

Table A.1: Summary statistics, by whether the school closure data is known

	Date unknown		Date known		Differences				
VARIABLES	mean	N	mean	N	Raw	p-value	+controls	p-value	Std. Diff.
Panel A: Socio-economic characteristics									
City (0/1)	0.013	299	0.062	1298	-0.049	0.001	0.000	.	-0.259
Population in 1917	1412.706	299	2970.103	1298	-1.6e+03	0.044	0.000	.	-0.164
Municipal tax rate 1917	35.246	299	12.782	1285	22.463	0.211	21.874	0.227	0.061
Local tax rate 1917	3.772	298	6.914	1294	-3.142	0.632	-3.241	0.623	-0.039
Factories/1000 capita	0.027	299	0.186	1298	-0.159	0.001	-0.018	0.365	-0.266
Factory workers/1000 capita	0.151	299	5.999	1298	-5.848	0.001	-1.240	0.239	-0.283
Labor and capital income/capita	247.989	299	321.289	1298	-73.300	0.002	-33.973	0.116	-0.209
Total production value/capita	1.710	299	63.772	1298	-62.062	0.001	-12.407	0.292	-0.276
Revenue/capita	19.282	299	23.496	1298	-4.214	0.000	-1.181	0.170	-0.262
Expenditure/capita	20.202	299	24.810	1298	-4.608	0.001	-1.250	0.196	-0.255
Assets/capita	48.086	299	69.444	1298	-21.358	0.000	-4.127	0.217	-0.277
Debt/capita	17.974	299	33.696	1298	-15.722	0.000	-4.103	0.062	-0.306
Recipients of poverty relief/1000 capita	30.085	299	30.495	1298	-0.410	0.696	0.929	0.345	-0.025
Poorhouses/1000 capita	13.463	299	12.389	1298	1.073	0.147	0.836	0.258	0.091
Military camp (0/1)	0.003	299	0.021	1298	-0.017	0.038	0.000	0.955	-0.160
Train station (0/1)	0.154	299	0.235	1298	-0.081	0.002	-0.073	0.006	-0.206
Panel B: 1917 national election shares									
Conservative party	0.287	290	0.288	1263	-0.001	0.948	-0.000	0.985	-0.004
Liberal party	0.299	290	0.289	1263	0.010	0.337	0.008	0.467	0.062
Labour party	0.234	290	0.279	1263	-0.045	0.000	-0.039	0.002	-0.230
Agricultural party	0.129	290	0.089	1263	0.040	0.000	0.037	0.000	0.243
Communist party	0.054	290	0.047	1263	0.007	0.361	0.008	0.282	0.056
Panel C: Health care and mortality									
Hospital (0/1)	0.007	299	0.037	1298	-0.030	0.007	-0.005	0.561	-0.208
Infant mortality rate 1916	61.561	295	66.081	1287	-4.520	0.301	-4.559	0.299	-0.065
Nr. of physicians per 1000	0.135	299	0.145	1295	-0.010	0.433	0.013	0.252	-0.050
Nr. of nurses per 1000	0.161	299	0.157	1295	0.004	0.602	0.010	0.199	0.032

NOTE.— This table reports summary statistics at the local level for pre-epidemic characteristics for all closure districts by whether the exact closure data is know or not. In our event study regressions only districts with know dates can be included. Otherwise the structure is the same as in Table 1.

Table A.2: Summary statistics, by speed of school closures, <5k population and excl. cities

	Slow closures		Fast closures		Differences				
VARIABLES	mean	N	mean	N	Raw	p-value	+controls	p-value	Std. Diff.
Panel A: Socio-economic characteristics									
City (0/1)	0.000	434	0.000	530	0.000	.	0.000	.	.
Population in 1917	1641.915	434	1421.349	530	220.566	0.001	0.000	.	0.217
Municipal tax rate 1917	24.344	433	6.999	529	17.344	0.261	18.612	0.230	0.069
Local tax rate 1917	13.234	433	3.576	529	9.658	0.257	10.210	0.234	0.070
Factories/1000 capita	0.000	434	0.000	530	0.000	.	0.000	.	.
Factory workers/1000 capita	0.000	434	0.000	530	0.000	.	0.000	.	.
Labor and capital income/capita	261.146	434	244.683	530	16.463	0.278	-0.755	0.958	0.070
Total production value/capita	0.000	434	0.000	530	0.000	.	0.000	.	.
Revenue/capita	20.413	434	19.543	530	0.870	0.336	0.878	0.335	0.062
Expenditure/capita	21.038	434	20.260	530	0.778	0.415	0.785	0.414	0.053
Assets/capita	47.431	434	47.937	530	-0.507	0.819	-0.734	0.742	-0.015
Debt/capita	18.559	434	20.851	530	-2.292	0.099	-2.836	0.041	-0.108
Recipients of poverty relief/1000 capita	29.143	434	27.709	530	1.434	0.110	0.549	0.524	0.103
Poorhouses/1000 capita	12.702	434	13.429	530	-0.728	0.352	-0.463	0.554	-0.060
Military camp (0/1)	0.000	434	0.002	530	-0.002	0.366	-0.002	0.320	-0.061
Train station (0/1)	0.217	434	0.198	530	0.018	0.481	-0.004	0.873	0.046
Panel B: 1917 national election shares									
Conservative party	0.273	426	0.305	517	-0.032	0.021	-0.024	0.083	-0.152
Liberal party	0.298	426	0.289	517	0.010	0.375	0.010	0.357	0.058
Labour party	0.282	426	0.272	517	0.010	0.409	0.003	0.786	0.054
Agricultural party	0.100	426	0.089	517	0.010	0.269	0.013	0.171	0.072
Communist party	0.047	426	0.033	517	0.014	0.022	0.010	0.108	0.148
Panel C: Health care and mortality									
Hospital (0/1)	0.009	434	0.006	530	0.004	0.518	0.002	0.654	0.041
Infant mortality rate 1916	64.808	433	64.690	527	0.118	0.978	0.207	0.961	0.002
Nr. of physicians per 1000	0.122	434	0.108	530	0.014	0.078	0.013	0.105	0.114
Nr. of nurses per 1000	0.165	434	0.140	530	0.024	0.001	0.021	0.004	0.221
All-cause mortality rate, 1917	14.131	434	14.693	530	-0.563	0.096	-0.516	0.129	-0.109
Days school closed	16.270	434	19.242	530	-2.972	0.000	-3.188	0.000	-0.243

NOTE.— This table reports summary statistics at the local level for pre-epidemic characteristics by response time, but only for our rural sample of school districts. Otherwise the structure is the same as in Table 2.

Table A.3: Long-run robustness checks: men, close vs not close

	DiDisc					
	$\pm 2y$	$\pm 1y$	$\pm 6m$	$\pm 1y$	$\pm 1y$ dnut	$\pm 6m$
Died before age 5	0.000	0.001	0.002	0.002	0.001	0.001
	(0.002)	(0.002)	(0.003)	(0.005)	(0.005)	(0.006)
N	209670	105183	52791	105183	101792	52791
Placebo death -1yr	-0.000	0.001	0.001	0.002	0.003*	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
N	209670	105183	52791	105183	101792	52791
Died during pandemic	-0.001	-0.001	0.000	0.000	0.002	-0.003
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
N	209670	105183	52791	105183	101792	52791
Age at death (years)	-0.252	-0.389	-0.704*	-0.724	-0.601	-0.779
	(0.225)	(0.306)	(0.422)	(0.597)	(0.626)	(0.844)
N	209666	105181	52790	105181	101790	52790
Responded Census 1950	-0.001	-0.001	-0.003	-0.003	-0.001	-0.007
	(0.003)	(0.005)	(0.006)	(0.009)	(0.010)	(0.013)
N	212566	106563	53475	106563	103131	53475
Employment 1950	-0.000	0.000	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
N	183132	92015	46169	92015	89068	46169
HISCAM 1950 (std.)	0.010	0.004	0.005	-0.003	0.003	-0.013
	(0.010)	(0.014)	(0.021)	(0.027)	(0.028)	(0.040)
N	171062	85968	43150	85968	83200	43150
Responded Census 1960	-0.002	-0.002	0.002	0.004	0.003	-0.005
	(0.004)	(0.005)	(0.007)	(0.009)	(0.010)	(0.013)
N	212566	106563	53475	106563	103131	53475
A levels	0.000	0.001	-0.000	-0.005	-0.005	-0.010
	(0.002)	(0.003)	(0.004)	(0.006)	(0.006)	(0.009)
N	180141	90495	45457	90495	87598	45457
Employment 1960	0.002	0.000	-0.001	-0.002	-0.001	-0.012*
	(0.002)	(0.003)	(0.004)	(0.005)	(0.006)	(0.008)
N	180141	90495	45457	90495	87598	45457
Responded Census 1970	-0.003	-0.004	-0.004	-0.003	-0.002	-0.008
	(0.004)	(0.006)	(0.008)	(0.010)	(0.011)	(0.015)
N	212566	106563	53475	106563	103131	53475
Sec. education 1970	-0.001	-0.002	-0.004	-0.008	-0.009	-0.013
	(0.003)	(0.004)	(0.006)	(0.008)	(0.008)	(0.011)
N	126390	84165	42298	84165	81486	42298
Employment 1970	0.001	-0.005	-0.010	-0.016	-0.022*	-0.021
	(0.004)	(0.005)	(0.008)	(0.012)	(0.012)	(0.016)
N	167829	84264	42347	84264	81581	42347
Retired 1970	-0.001	0.004	0.009	0.014	0.016	0.018
	(0.004)	(0.005)	(0.007)	(0.010)	(0.010)	(0.014)
N	167829	84264	42347	84264	81581	42347
Total Income 1970 all	168.820	68.615	-302.241	-559.700	-614.653	-1821.233*
	(211.708)	(331.268)	(464.996)	(639.282)	(677.240)	(961.852)
N	167829	84264	42347	84264	81581	42347
Total Income 1970 all (ln)	-0.002	-0.010	-0.018	-0.015	-0.015	-0.036
	(0.007)	(0.010)	(0.014)	(0.020)	(0.021)	(0.028)
N	155759	78220	39312	78220	75727	39312

Notes: This table reports the coefficients from different difference-in-differences and difference-in-discontinuities estimations. Each coefficient represents one regression. The DiD regressions uses children born in different windows between 1910 and 1913 and fixed effects for parish, birth year and birth month (see equation 3; standard errors are clustered at the parish level. The DiDisc regressions uses children born in different windows between 1911 and 1912 according to equation 3; dnut indicates doughnut specification which leaves out children born in the two weeks surrounding January 1st; standard errors are clustered at the running variable, i.e. day of birth.

Table A.4: Long-run robustness checks: women, close vs not close

	DiD			DiDisc		
	$\pm 2y$	$\pm 1y$	$\pm 6m$	$\pm 1y$	$\pm 1y$ dnut	$\pm 6m$
Died before age 5	-0.001	0.001	-0.002	-0.003	-0.004	0.001
	(0.002)	(0.002)	(0.003)	(0.005)	(0.005)	(0.007)
N	200001	100258	50679	100258	96944	50679
Placebo death -1yr	-0.001	-0.001	-0.000	0.000	-0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
N	200001	100258	50679	100258	96944	50679
Died during pandemic	-0.000	0.001	0.002	0.002	0.000	-0.003
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)
N	200001	100258	50679	100258	96944	50679
Age at death (years)	-0.178	-0.592*	-0.336	-0.361	-0.063	-0.281
	(0.232)	(0.319)	(0.475)	(0.675)	(0.699)	(0.973)
N	200000	100257	50679	100257	96943	50679
Responded Census 1950	-0.004	-0.006	0.000	-0.001	0.004	0.005
	(0.003)	(0.004)	(0.007)	(0.009)	(0.010)	(0.014)
N	203675	101997	51521	101997	98610	51521
Employment 1950	-0.001	0.003	-0.002	-0.005	-0.005	-0.005
	(0.003)	(0.005)	(0.007)	(0.009)	(0.010)	(0.013)
N	177510	89007	45060	89007	86058	45060
HISCAM 1950 (std.)	-0.031	-0.011	-0.028	-0.023	-0.040	-0.056
	(0.021)	(0.029)	(0.043)	(0.062)	(0.064)	(0.098)
N	40527	19931	9829	19931	19290	9829
Responded Census 1960	-0.003	-0.004	0.000	-0.002	0.002	0.009
	(0.003)	(0.005)	(0.007)	(0.010)	(0.010)	(0.014)
N	203675	101997	51521	101997	98610	51521
A levels	-0.001	-0.001	0.000	-0.000	0.000	-0.002
	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)
N	175071	87703	44409	87703	84810	44409
Employment 1960	0.005	0.001	0.002	0.003	0.005	0.028
	(0.005)	(0.007)	(0.011)	(0.014)	(0.015)	(0.021)
N	175071	87703	44409	87703	84810	44409
Responded Census 1970	-0.005	-0.009*	-0.005	-0.007	-0.001	-0.004
	(0.004)	(0.005)	(0.008)	(0.011)	(0.011)	(0.016)
N	203675	101997	51521	101997	98610	51521
Sec. education 1970	-0.001	-0.003	-0.004	-0.002	-0.002	-0.006
	(0.003)	(0.004)	(0.005)	(0.007)	(0.008)	(0.010)
N	125940	83870	42430	83870	81100	42430
Employment 1970	0.014***	0.019***	0.023**	0.033**	0.026*	0.058***
	(0.005)	(0.007)	(0.009)	(0.013)	(0.014)	(0.019)
N	167422	83922	42462	83922	81151	42462
Retired 1970	-0.011**	-0.022***	-0.017	-0.023	-0.022	-0.053**
	(0.005)	(0.008)	(0.011)	(0.016)	(0.017)	(0.023)
N	167422	83922	42462	83922	81151	42462
Total Income 1970 all	197.574*	301.541*	186.716	306.800	381.222	945.632*
	(117.894)	(167.120)	(218.146)	(338.202)	(349.135)	(493.049)
N	167422	83922	42462	83922	81151	42462
Total Income 1970 all (ln)	0.003	-0.005	0.035	0.057	0.076*	0.060
	(0.014)	(0.021)	(0.030)	(0.043)	(0.043)	(0.064)
N	101776	50280	25287	50280	48613	25287

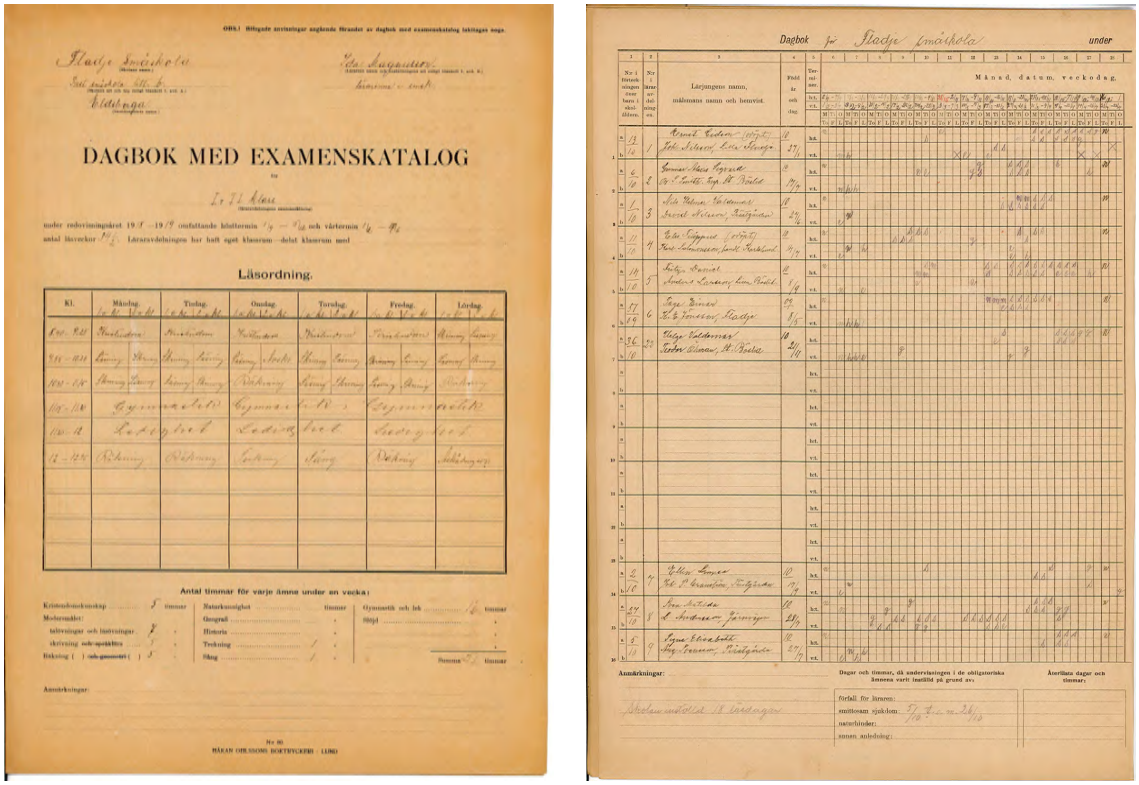
Notes: This table reports the coefficients from different difference-in-differences and difference-in-discontinuities estimations. Each coefficient represents one regression. The DiD regressions uses children born in different windows between 1910 and 1913 and fixed effects for parish, birth year and birth month (see equation 3; standard errors are clustered at the parish level. The DiDisc regressions uses children born in different windows between 1911 and 1912 according to equation 3; dnut indicates doughnut specification which leaves out children born in the two weeks surrounding January 1st; standard errors are clustered at the running variable, i.e. day of birth.

Table A.5: Long-run estimates of school closures, grades 3/4

	Men		Women	
	DiD	DiDisc	DiD	DiDisc
Panel A: Mortality outcomes				
Died before age 5	0.002	0.012**	-0.001	-0.004
	(0.002)	(0.005)	(0.002)	(0.005)
Died during pandemic	0.001	0.002	0.001	0.005***
	(0.001)	(0.002)	(0.001)	(0.002)
Age at death (years)	-0.181	-0.105	-0.293	0.115
	(0.221)	(0.667)	(0.238)	(0.709)
N	211689	106946	204509	103363
Panel B: 1950 Census				
Responded Census 1950	-0.006*	-0.004	-0.005	-0.006
	(0.004)	(0.010)	(0.003)	(0.010)
N	214240	108238	207115	104645
Employment 1950	0.000	-0.001	0.001	-0.001
	(0.001)	(0.003)	(0.003)	(0.010)
N	181274	91683	177648	89809
HISCAM 1950 (std.)	-0.010	0.005	-0.020	-0.027
	(0.010)	(0.028)	(0.021)	(0.058)
N	169311	85578	42546	21295
Panel C: 1960 Census				
Responded Census 1960	-0.003	0.007	-0.003	-0.002
	(0.004)	(0.011)	(0.004)	(0.011)
N	214240	108238	207115	104645
A levels	0.001	0.005	0.001	0.000
	(0.002)	(0.006)	(0.001)	(0.003)
N	176795	89495	174027	87996
Employment 1960	-0.000	0.002	-0.001	0.004
	(0.002)	(0.007)	(0.005)	(0.015)
N	176795	89495	174027	87996
Panel D: 1970 Census				
Responded Census 1970	-0.003	0.010	-0.004	-0.001
	(0.004)	(0.012)	(0.004)	(0.012)
N	214240	108238	207115	104645
Employment 1970	0.004	0.005	-0.008*	-0.013
	(0.005)	(0.014)	(0.005)	(0.012)
N	160632	81342	164021	82924
Retired 1970	-0.001	-0.002	0.005	-0.015
	(0.005)	(0.013)	(0.005)	(0.015)
N	160632	81342	164021	82924
Total Income 1970 all	53.890	241.461	-14.762	-133.313
	(206.821)	(584.961)	(121.810)	(305.228)
N	160632	81342	164021	82924

Notes: This table reports the coefficients from the difference-in-differences and difference-in-discontinuities estimations for different outcomes. Each coefficient represents one regression. The DiD regressions uses children born between 1907 and 1910 and fixed effects for parish, birth year and birth month (see Equation 3); standard errors are clustered at the parish level. The DiDisc regressions uses children born between 1908 and 1909 according to Equation 3; standard errors are clustered at the running variable, i.e. day of birth.

Appendix B Main Sources



DAGBOK MED EXAMENSKATALOG

Läsordning.

Dagbok för Gladje Småskola under

Figure B.1: Exam Catalogue from Elementary School

Note.— The extract shows the front page and the critical page of an exam catalog from the school year 1918/19.

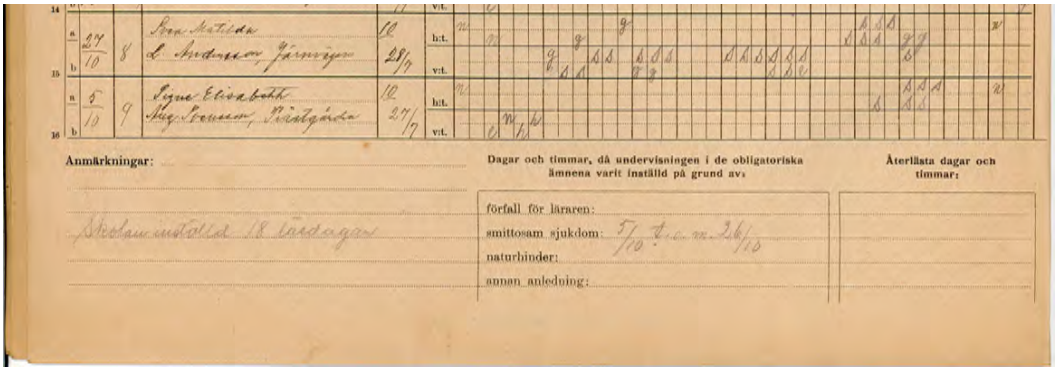


Figure B.2: Notes on School Closures

Note.— The extract shows the critical piece of information contained in the exam catalogue regarding school closures. In the box, the class teacher puts down instances of interruptions to normal operation by cause, including absence of the teacher, epidemic, natural causes and other causes.

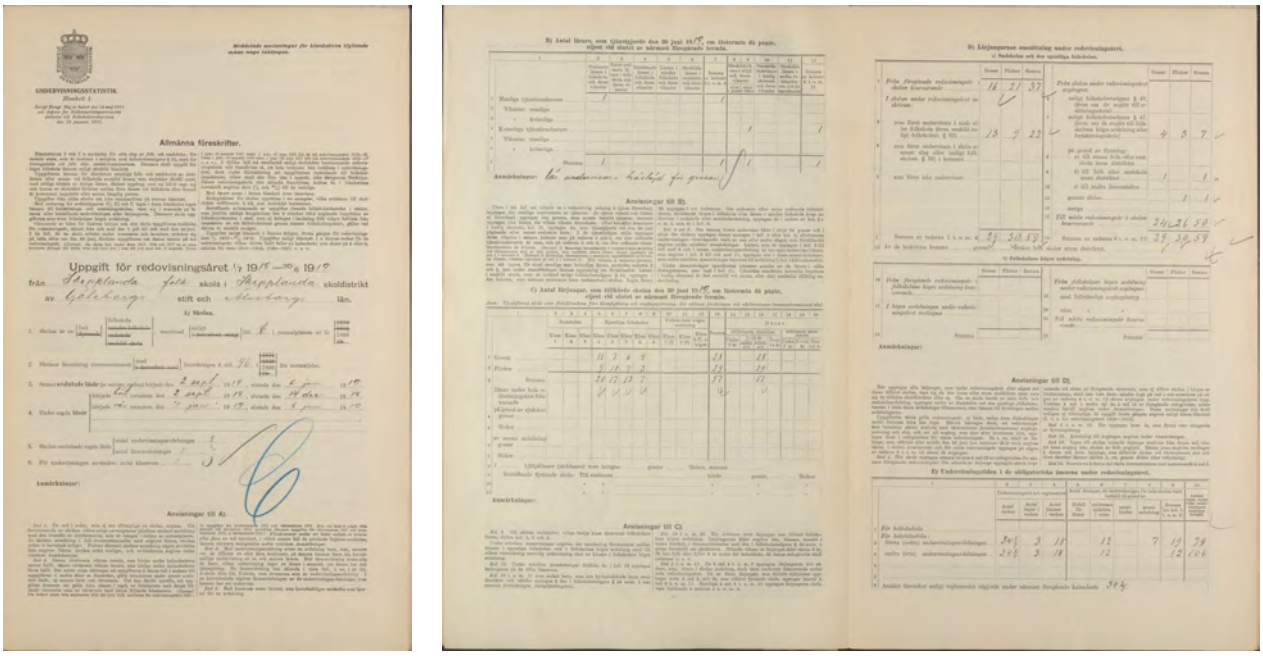
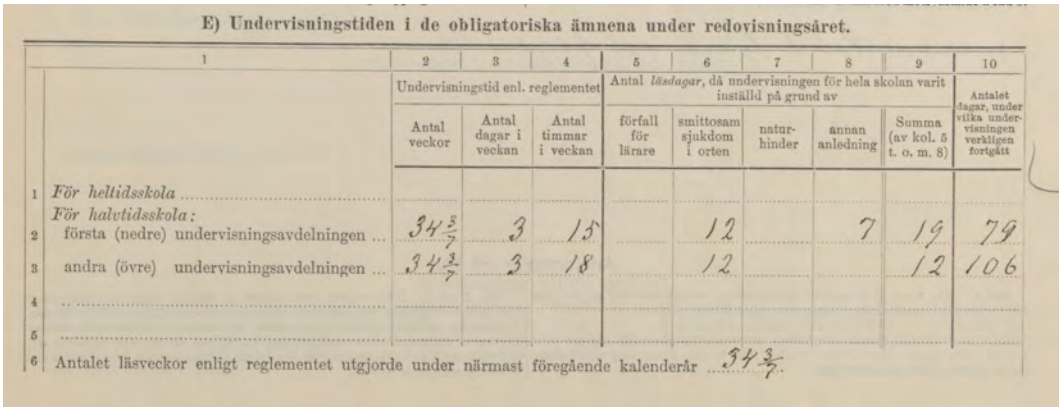


Figure B.3: Alternative Source: Teaching Statistics

NOTE.— The extract shows three pages of the teaching statistics (*Undervisningsstatistik*) form that was completed for each school in the country on an annual basis.



E) Undervisningstiden i de obligatoriska ämnena under redovisningsåret.

1	2	3	4	5	6	7	8	9	10
	Undervisningstid enl. reglementet			Antal *läsdagar*, då undervisningen för hela skolan varit inställd på grund av					Antalet dagar, under vilka undervisningen verkligen fortgått
	Antal veckor	Antal dagar i veckan	Antal timmar i veckan	förfall för lärare	smittosam sjukdom i orten	natur-hinder	annan anledning	Summa (av kol. 5 t. o. m. 8)	
1 *För heltidsskola*									
För halvtidsskola:									
2 första (nedre) undervisningsavdelningen ...	34 3/7	3	15		12		7	19	79
3 andra (övre) undervisningsavdelningen ...	34 3/7	3	18		12			12	106
4									
5									
6 Antalet läsveckor enligt reglementet utgjorde under närmast föregående kalenderår ...	34 3/7								

Figure B.4: Alternative Source: Teaching Statistics

NOTE.— The extract shows the part of the teaching statistics form that reports school closures. In the box, the head teacher puts down the number of weeks of instruction; the number of days of instruction per week; days lost due to teacher absence, epidemics, natural causes and other causes; and the total number of days of instruction during the academic year. In many cases, the exact dates of school closures were added on the following page.

Appendix C Data Digitization

This section describes how to use the end-to-end automated transcription pipeline of handwritten text proposed by Dahl et al. (2023) on the Swedish death books. Every component of this pipeline is optimized to handle large amounts of high resolution scans (pdfs or images) with a fixed tabular layout. The pipeline is particularly well suited to transcribe handwritten population data such as dates, cause-of-death, occupation, gender, civil status etc., that are organized in well-structured tables as is the case for the Swedish death books.

C.1 The transcription pipeline

Following Dahl et al. (2021) the transcription workflow can be divided into three steps:

1. Layout classification: Sort documents based on type.
2. Table Segmentation: Extract an image for each of the fields of interest.
3. Transcription: Transcribe the extracted field images.

1. Layout classification:

For the Swedish death books downloaded from riksarkivet.se there are 2 fixed layout types. We define Type A as having a table layout with 11 entry-rows while Type B is a table layout with 12 entry-rows, see, e.g., Figure C.1. The classification of all documents into layout types is based on a convolutional neural network. The layout classifier performs with an accuracy rate above 99%.

Table segmentation:

For table segmentation, firstly, the exact location of the table in each image document is identified. Secondly, all the table cells of interest are extracted and stored as smaller

image-snippets. Table segmentation is based on standard computer vision operations to find straight horizontal and vertical lines and their points of intersection. The Coherent Point Drift (CPD) method of Myronenko and Song (2010) can then be applied to learn the transformation responsible for aligning the points of intersection on a table template with the points found on the image document. Based on the learned transformation, all the table cells of interest can subsequently be extracted, see, e.g., Figure C.2. For more details, see Dahl et al. (2023). For the Swedish death books, more than 95% of all tables where successfully segmented.

Transcription:

Having extracted a collection of image-snippets corresponding to the fields of interest in the tables, we need them transcribed. This is the process of converting an image snippet of numbers and text into a sequence or string representation. For all image-snippets extracted from the death books we use a variant of a ResNet-50 sequence neural network. Specifically, each neural network uses a ResNet-50 with bottleneck building blocks (He et al., 2016) as its

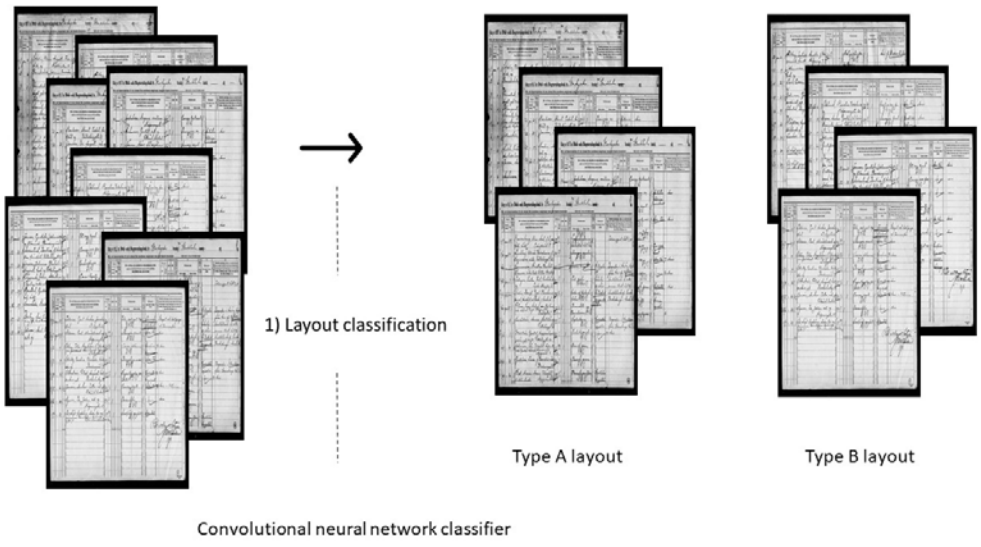
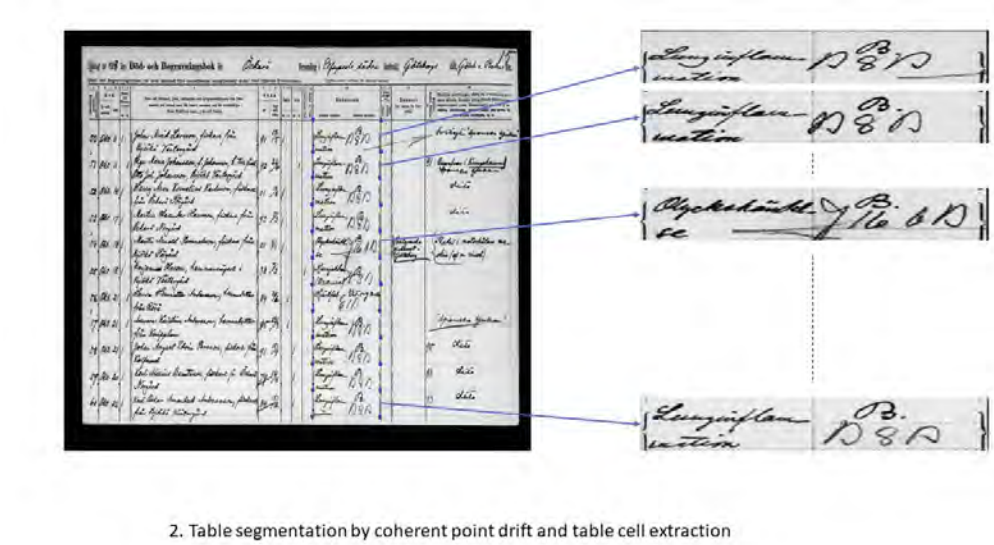


Figure C.1: Table layout classification by a convolutional neural network



2. Table segmentation by coherent point drift and table cell extraction

Figure C.2: Table segmentation by coherent point drift and extraction of table cells of interest

feature extractor; the weights of the PyTorch version of ResNet-50 pretrained on ImageNet (Deng et al., 2009) are used as the initial weights. This step is illustrated in Figure C.3 for cause-of-death. For dates and civil status, we obtain accuracy rates between 90% and 95%. For cause-of-death, the accuracy is about 92%.

Neural network architecture: The architecture of the neural networks models differ across the various table cells only insofar as their classification heads differ. The neural network for modeling sequences of the format of DD-M-YY (the date-of-birth cells, i.e., columns 7 and 8 of the death book) have 5 layers. The first two correspond to the day, with four and ten nodes, the third to the month, with fourteen nodes and the last two to the year, with 11 nodes each. The thirteenth option for the month allows for missing months and the fourteenths option for the month allows for the "repeat" option. The eleventh option for each digit in the year allows for missing digits. The architecture of the network modeling sequences of the format DD-M (the date-of-death cells, i.e., columns 2 and 3 of the death book) is defined similarly except for the last two layers which are dropped. The network

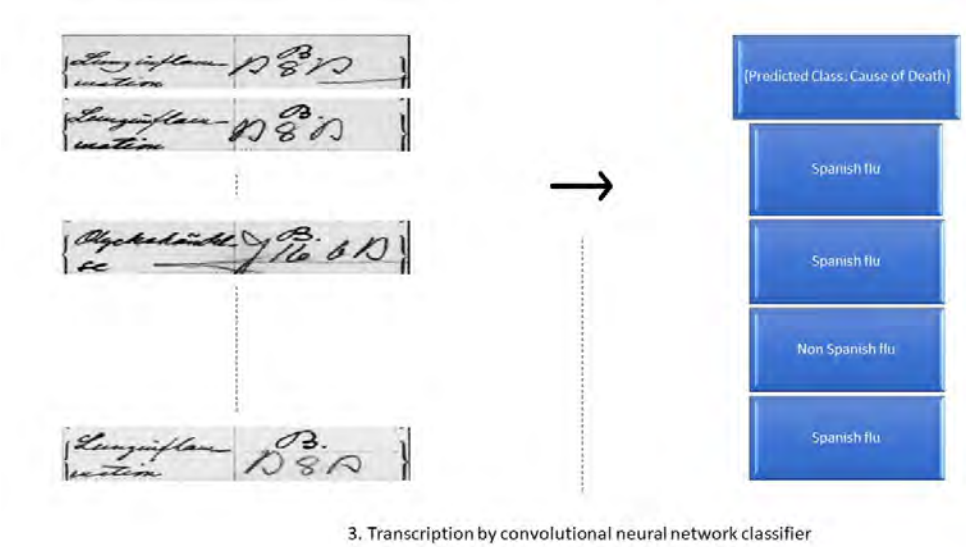


Figure C.3: Transcription of cause-of-death class by a convolutional neural network

for modeling the cause-of-death (column 16) is a single layer classifier with two nodes corresponding to Spanish flu versus non Spanish flu respectively. Each row from columns 9 to 15 is segmented at a single image snippet and is transcribed by a single layer neural network with 7 nodes where the first node is for an empty image snippet (0-0-0-0-0-0-0), the second node is a "mark" in column 9 (1-0-0-0-0-0-0) etc.

Neural network configuration and optimization: All neural networks are optimized using stochastic gradient descent with momentum of 0.9, weight decay of 0.0005, and Nesterov acceleration based on the formula in Sutskever et al. (2013). The batch size used is 256 and the learning rate is 0.05. The networks are trained for 100 epochs and the learning rate is divided by ten every 30 epochs. The loss function used is the mean cross entropy of each output layer.

C.2 Training data

The Swedish death books for the period 1916-2020 are already partly transcribed in the sense that Swedish administrative death records contain information about names, birth and death dates, gender, civil status etc. Furthermore, when retrieving the source images from riksarkivet.se, we automatically obtain meta data information on the parish for each death book image as well as the year of death for all individuals in that particular death book. Surprisingly, the cause-of-death is **not** recorded in the Swedish administrative death records for the 1916-2020 period time period. Furthermore, the entries of Swedish administrative death records are not linked to the death books. In order to take advantage of the large amount of transcribed but unlinked administrative data we apply a two-step transcription procedure:

Dates and Civil Status: First training based on manually transcribed labels.

We first train two initial networks based on subsets of 4000 manually labeled death dates images and 8000 labeled birth date images. Based on predictions from these initial network models for birth date and death date (in combination with the meta data on parishes) we are able to perfectly match 160,000 individuals from the Swedish administrative records to the death book images in which the individuals are recorded. This implies that we after this first round have labels for about 160,000 image snippets for birth dates, death dates, gender and civil status. All this information is also linked to the unique death book image identifier, i.e., the filename of the image at Riksarkivet.se.

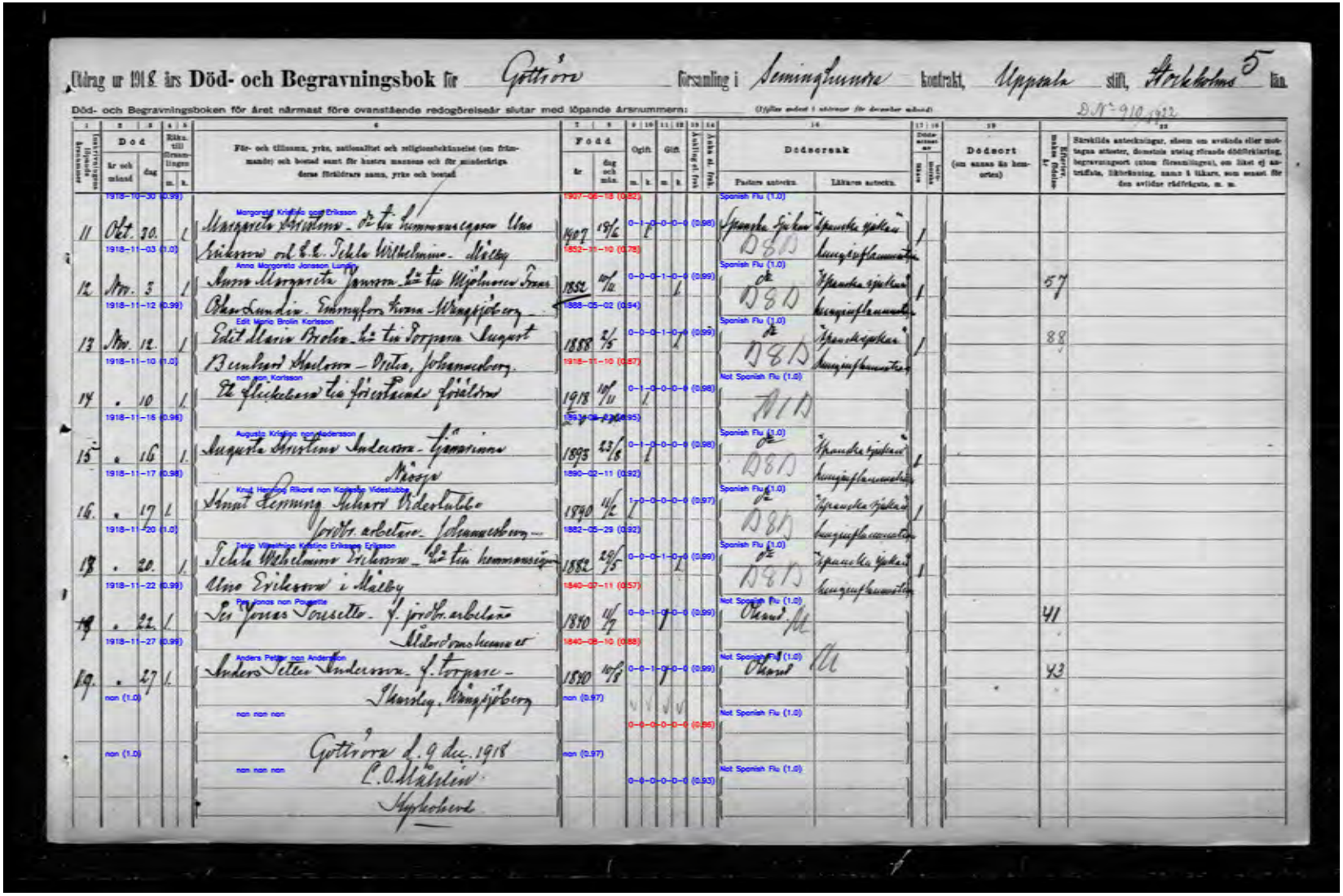
Dates and Civil Status: Second training based on linked labels.

Consequently, in the second step we retrain the networks for the dates as well as civil status based on this much larger training data and use these models to automatically transcribe all remaining image snippets.

Cause-of-death: Training based on manually transcribed labels.

Finally, the neural network classification model for cause-of-death is trained based on 10000 manually labeled cause-of-death image snippets. Based on this model the cause of death is predicted for all the remaining cause of death image snippets and the predictions are linked by filename and row number to the already obtained transcriptions for each individual.

Figure C.4: Example of a fully transcribed Swedish death book



NOTE.— The transcription in a given cell is written in blue if the transcription confidence (the figure in parenthesis) is above 90 percent otherwise the transcription is written in red. The names in column 6 are not transcribed but are used for visual inspection of correct/incorrect matching to the official administrative Swedish death records. If the name written in blue equals the handwritten name in the associated cell the matching is correct.

Appendix D Pupil Dataset – Descriptives and Results

D.1 Data Description

We digitised and linked a subset of exam catalogues (cf. Section 3.1) to censuses. The sample consists of more than 16,000 individuals from 362 school districts. The sample is based on catalogues that have been sent by archives and the selection of cases is thus arbitrary but may still not be completely representative due to the fact that we oversample larger schools. In addition, some archives only sent the first page of the list of pupils of that school, and since boys are typically reported before girls, this may lead to some overrepresentation of boys. It is, however, unlikely that the sample selection is somehow related to relevant unobservable characteristics. We provide an overview of the descriptives of the resulting dataset in Table D.1.

Table D.1: Summary Statistics, Pupil Sample

	Mean	S.d.	Min	Max	*N*
Absence days fall	4.63	6.70	0.0	115.0	16,453
Absence days spring	3.58	7.21	0.0	108.0	16,448
Presence days fall	78.50	47.00	0.0	4775.0	16,449
Presence days spring	54.42	21.31	0.0	434.0	16,439
Presence rate (%)	94.32	6.75	21.1	100.0	16,115
Presence rate (%) fall	92.37	10.27	3.8	100.0	16,249
Presence rate (%) spring	95.79	7.65	1.1	100.0	16,309
Half weak reading	0.22	0.42	0.0	1.0	16,468
Half year reading	0.09	0.28	0.0	1.0	16,468
Closure days	13.68	13.24	0.0	88.0	16,264
No closure	0.29	0.45	0.0	1.0	16,264
Distance from POB (km)	54.85	133.31	0.0	1376.0	13,097
Measurement error birth/18	-0.50	11.22	-51.0	51.0	12,046

NOTE.— This table reports summary statistics for pupils in a non-random sample of exam catalogues from 362 school districts. *Distance from POB* refers to the distance between the 1918 school district and the parish of birth. *Measurement error birth/18* refers to the measurement error (in school closure days) when parish of birth is used for treatment assignment instead of the actual school district.

For each individual pupil, the data include information of the total days of absence and

the total days of presence in each term of the 1918/19 school year. It also includes information on the school regulations applying at the individual school; most importantly, whether half-time schooling was applied. Some schools had one teacher teaching two groups; the groups would take turns either on individual days within each week (denoted "Half week reading" in Table D.1) or be taught one full term each (denoted "Half year reading" in Table D.1). Thanks to record linkage to the 1910 census, the dataset also includes parish of birth and parish of residence in 1910. Thereby, it is also possible to calculate the measurement error if the assignment of the school closure variable is based on place of birth instead of the actual exposure in 1918 – which is the treatment assignment we use in the analysis of long-term effects in Section 5.

D.2 Closures, Presence and Sickness Absence

In Table D.2, we regress student presence on days of closure, in order to get an estimate of the rate at which school closures are translated into reduced presence. We consider three different assignments of the school closure assignment variable. In Panel A, we use the days of school closure recorded for the district where the pupil actually went to school in 1918/19. In Panel B, we instead base the assignment on the parish of birth. The three leftmost columns of Table D.2 report the effect of closure days on the number of days a student is present in school, whereas the three rightmost columns report the corresponding effects on presence *rates* (calculated as the number of presence days divided by the combined number of presence and sickness absence days).

The overall effect of a day of school closure is a reduction in presence days by 0.45. There are a number of reasons why this coefficient is not equal to 1. First, we have defined school closures by calendar dates, so that two weeks of closure counts as 14 days – even though there were only 5 or 6 schooldays per week. Second, the counterfactual to a school closure is not a 100% presence, since pupils may be absent even if their school is open, due

to e.g. sickness of the pupils or their teacher. Third, there is sometimes measurement error in the school closure variable: we take school closure information from the *largest* school in each district, and some districts have variation between schools. Fourth, on rare occasions, there would be an extension to the term due to the closure; however, this represents a rare exception.

Despite these deviations, the estimate in the first column makes clear that there is a strong negative relationship between closures and presence in school. The following columns make clear that, as expected, it is mainly presence during the fall term that is affected. In the three rightmost columns, we estimate the effect on the presence *rates*. The rate is calculated as the total days of presence divided by the total days of instruction. Therefore, finding an effect on this variable – which by construction discounts the period of school closure – would indicate systematic differences in presence between closing and non-closing districts also in normal times, which in turn would reflect an endogeneity problem for the school closures. Interestingly, there is a **positive** relationship between the duration of a school closure and the presence rates. This means that the schools that closed actually had higher presence rates before and after the closure. The coefficient is small: each day of closure would be associated in a reduction in the absence rate by 0.3 per cent. This is one reason why each day of closure doesn't translate into one day less of presence; however, it is clearly not quantitatively important.

In panel B of Table D.2, we estimate the impact of closures on presence days and rates when we do not use the actual exposure to closures, but instead assign closures by the parish of birth. This leads to a bias in coefficients toward zero; however, the key relationship remains strongly statistically significant. We will return to the issue of measurement error below.

In Figure D.1, we further investigate the relationship between school closures and presence rates during times when schools are open. It shows the average presence rates

Table D.2: Effects of Closures on Presence

	Presence Days			Presence Rates		
	Total	Fall	Spring	Total	Fall	Spring
A. By Actual Exposure						
Days closed	-0.4490***	-0.3912***	-0.0500	0.0325*	0.0321	0.0193
	(0.126)	(0.072)	(0.093)	(0.013)	(0.027)	(0.011)
Mean	133.82	54.68	78.64	94.30	92.27	95.81
N. of cases	12,493	12,696	12,691	12,488	12,551	12,625
N. of districts	402	403	403	402	403	403
B. Exposure Measured at Birth						
Days closed	-0.4192***	-0.2895***	-0.1137	0.0187*	0.0198	0.0077
	(0.093)	(0.053)	(0.059)	(0.009)	(0.017)	(0.009)
Mean	133.82	54.68	78.64	94.30	92.27	95.81
N. of cases	9,984	10,151	10,147	9,981	10,023	10,103
N. of districts	513	515	516	513	514	516

Note.— *** p <0.01; ** p <0.05; * p <0.1. Own calculations based on pupil sample. Standard errors are clustered at the school district level. Covariates included in all specifications are year of birth fixed effects, term length, and dummies for half-time schooling.

during fall 1918 and spring 1919 for closing and non-closing schools. Apparently, the presence rates are very similar in closing and non-closing schools.

D.3 The Endogeneity of School Closures

Our baseline results reported in Section 4.2 suggest that school closures were typically implemented in reponse to a rapid upsurge in the number of fatal influenza cases. We now investigate whether similar patterns in pre-closure health may be detected in the sickness absence data. In the digitised data we have information on the number of sickness days per month at the individual level. The data are thus much more aggregated than the mortality data, which are available at the daily level – and this limits the possibilities somewhat.

In Table D.3 we present results showing how sickness absence rates change in the months leading up to a closure. The non-closing districts serve as reference category. In the

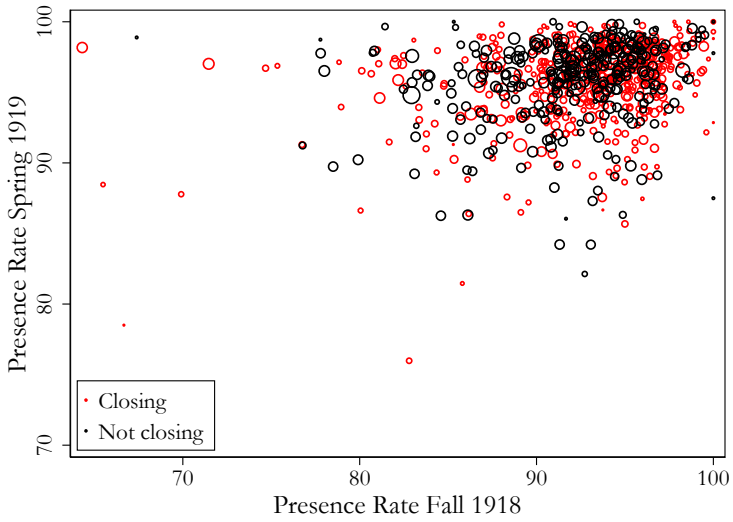


Figure D.1: Presence Rates in Closing and Non-Closing Schools

NOTE.— Own calculations based on pupil sample. Each marker represents a school district and shows presence rates (in per cent of total instruction time) in the fall of 1918 and the spring of 1919 for closing and non-closing districts. Districts are weighted by the number of pupils covered in the data.

three rightmost columns we test whether closures were preceded by diverging trends, using first-differenced absence rates as dependent variable. It transpires that the October closures are preceded by a surge in sickness absence, whereas all the other estimates are insignificant and often take on the "wrong" sign. Accordingly, schools closing in October had an 1.2 percentage point increase in absence in that month relative to the non-closing districts. In conclusion, we find that the October closures appear to have been endogenous in the sense that they were preceded by a surge in sickness absence: absence rates increase by more than 50 per cent in the pre-closure days of October. Hence in this part we corroborate the finding in 4.2 that closures were preceded by a surge in the pandemic.

D.4 Measurement Error in Exposure

Thanks to having place of residence measured at two points in time – at birth and in 1918 – we can gauge the extent to which measurement error in the assignment to school closures

Table D.3: Absence Rates by Closure Months

	Sickness absence rates in					Monthly Change		
	August	September	October	Nov/Dec	Spring 19	September	October	Nov/Dec
Closed September	0.0076	0.0112			-0.0033	-0.0015		
	(0.009)	(0.010)			(0.004)	(0.007)		
Closed October	-0.0009	-0.0090	0.0067		0.0032	-0.0043	0.0166**	
	(0.004)	(0.006)	(0.004)		(0.004)	(0.005)	(0.006)	
Closed Nov/Dec	0.0073	-0.0045	0.0054	0.0195	-0.0058	-0.0054	0.0088	0.0142
	(0.007)	(0.006)	(0.004)	(0.016)	(0.004)	(0.006)	(0.006)	(0.015)
Mean	0.030	0.030	0.025	0.104	0.045	-0.004	-0.003	0.084
N. of cases	6,430	11,232	10,161	6,085	12,020	6,422	10,015	6,031
N. of districts	240	378	350	219	392	240	344	219

Note.— *** p <0.01; ** p <0.05; * p <0.1. Standard errors are clustered at the school district level. Covariates included in all specifications are year of birth fixed effects, term length, and dummies for half-time schooling.

represents a problem. In Figure D.2 we plot histograms of the distance between the parish of residence in 1918 and parish of residence at birth. Clearly, the vast majority of individuals remain in a close vicinity of their earlier parish of residence: 70 per cent of the sample lives within 28 kilometers of their parish of birth. Of these, the vast majority remain in the same parish.

In figure D.3 we consider the implications of this for measurement error: we calculate the measurement error as the absolute value of the difference in total closure duration. Around 2/3 of the sample have a measurement error of two days or less.

Next, we consider the consequences of measurement error for the empirical analysis. Figure D.4 shows how the measurement error relates to the true value of the closure duration. In this figure we use the actual measurement error, i.e. the deviation from the true value, without taking absolute values. Not surprisingly, given that the variable is non-negative with great mass at zero closure days, the measurement error is clearly non-classical.

Given this strong correlation between measurement error and duration, we cannot make the assumption that the measurement error is uncorrelated with duration. By utilizing both the accurate and inaccurately measured exposure data within the pupil dataset, we are able to assess the degree of bias stemming from measurement errors in our primary model

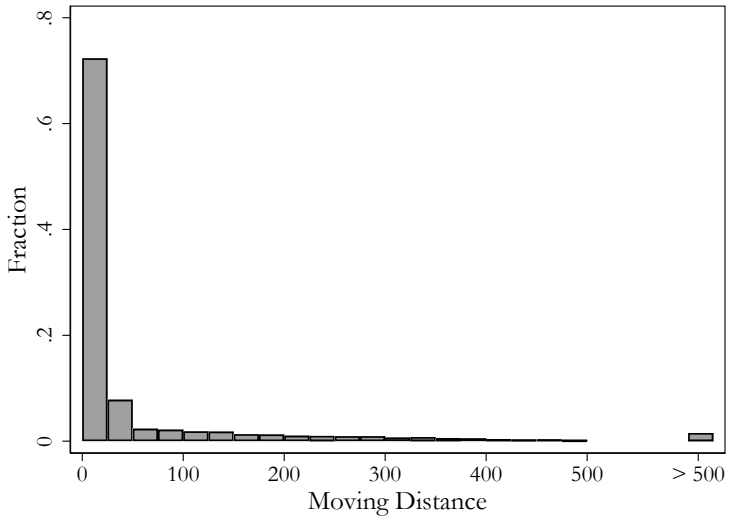


Figure D.2: Distance between 1918 Residence and Earlier Parish of Residence.

NOTE.— Own calculations based on pupil sample. Moving distance is measured as haversine distance between the 1918 school district centroid and parish of birth.

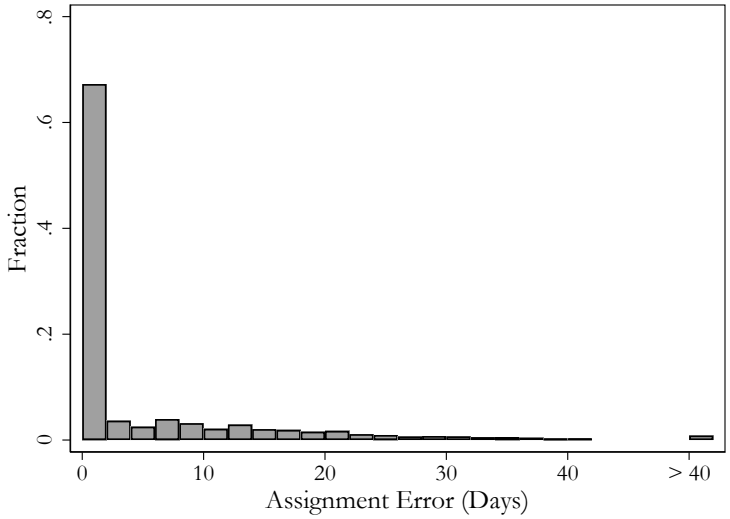


Figure D.3: Measurement Error in School Closure Duration.

NOTE.— Own calculations based on pupil sample. Measurement error is defined as the absolute value of the difference in closure duration between the 1918 school district and the parish of residence in 1910 or at birth.

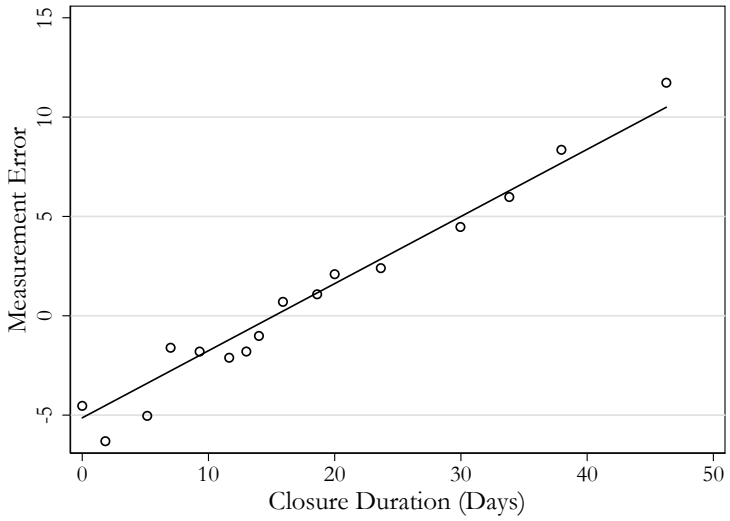


Figure D.4: Average Measurement Error by Closure Duration.

NOTE.— Own calculations based on pupil sample. Measurement error is defined as the difference in closure duration between the 1918 school district and the parish of birth.

for long-term outcomes.[25] Consider our main difference-in-difference specification from Section 5:

$$y_{impy} = \gamma_p + \gamma_m + \gamma_y + \gamma_D D_i + \tau C_p \cdot D_i + \epsilon_i, \tag{3}$$

which under the identifying assumptions produces unbiased estimates in case there is no measurement error in the independent variables. In fact, all variables in this specification are measured without error, except for the interaction term $C_p \cdot D_i$, which interacts the treatment status of the school district with the child's eligibility.[26] To simplify, we collect all the other, correctly measured variables, in the vector Z_i. Denote the mismeasured treatment assignment variable by $\widetilde{C}_p$ and the measurement error by $u_i = \widetilde{C}_p - C_p$. Then we may express

[25]The short-run anaysis for mortality is based on actual parish of residence at death and hence the issue of measurement error does not arise.

[26]In the main analysis, the relevant cutoff for D_i is January 1st 1912, since children born before that date went to school in 1918. In this part we use instead January 1st 1910 as the cutoff, since the entire sample we use in this part includes schoolchildren. This implies that the measurement error may be slightly overestimated, since it is based on slightly older children. However, it is unusual that children relocate during their primary school years (cf. Fischer et al., 2020); hence the implications of this should be negligible.

the estimating equation as:

$$y_{impy} = Z\gamma + \tau\widetilde{C}_p \cdot D_i + \epsilon_i - \tau u_i \cdot D_i \tag{4}$$

Following Aigner et al. (1973) it can be shown that the errors in the OLS estimates of parameters $(\gamma, \tau)'$ equal

$$e = -\tau \left(\widetilde{X}'\widetilde{X}\right)^{-1} \widetilde{X}' \left(u \odot D\right) \tag{5}$$

where $\widetilde{X} = \left(Z, \widetilde{C}_p \odot D\right)$ is a $N \times K$ matrix including all independent variables, including the mismeasured interaction term, and $\odot$ denotes element-by-element multiplication. Hence, the bias term may be estimated by regressing the vector $(u \odot D)$ on all the variables in $\widetilde{X}$. Denoting the coefficient associated with the term $\widetilde{C}_p \cdot D_i$ in such a regression by $b_{\widetilde{C}D}$, the bias associated with the DID specification equals factor $1 - b_{\widetilde{C}D}$. In other words, the parameter $\tilde{\tau}$ that results when running the difference-in-differences analysis with the mismeasured variable $\widetilde{C}_p$ is related to the true effect τ as follows:

$$\tilde{\tau} = \tau \left(1 - \left(\widetilde{X}'\widetilde{X}\right)^{-1} \widetilde{X}' \left(u \odot D\right)\right) \tag{6}$$

$$= \tau \left(1 - b_{\widetilde{C}D}\right) \tag{7}$$

In our main specification, the closure variables enters as a binary term, where $C_p = 1$ represents school closures of any length. However, the bias factor would be derived in the same way in a specification where it enters as a continuous variable. Therefore, for the sake of comparison, we estimate the bias factor for a binary and for a continuous specification in Table D.4 below.

As the estimates in the rightmost column of Table D.4 reveal, the measurement error in the school closure assignment leads to attenuation bias: the actual effect will be about 35 per cent larger than the estimated effect. Hence, all our estimates for long-term outcomes

will slightly underestimate the actual effect. It is notable that using a continuous treatment indicator instead would lead to a bias in the opposite direction.

Table D.4: Bias due to Measurement Error

	Continuous	Binary
Bias Factor $\left(1 - b_{\widetilde{CD}}\right)$	1.2760***	0.7420***
	(0.013)	(0.014)
N. of cases	6,040	6,040

*** p <0.01; ** p <0.05; * p <0.1.

It would in principle be possible to also estimate the bias resulting from measurement error in the difference-in-discontinuity specification (cf. equation (3) in Section 5). However, since that specification includes several mismeasured variables (i.e. all interactions with the treatment variable) it does not lead to a simple solution.

References for Online Appendix

Aigner, Dennis J et al., "Regression with a binary independent variable subject to errors of observation," *Journal of Econometrics*, 1973, *1* (1), 49–59.

Dahl, C., Torben S. D. Johansen, E. N. Sørensen, Christian Westermann, and Simon F. Wittrock, "Applications of Machine Learning in Document Digitisation," *ArXiv*, 2021, *abs/2102.03239.*

Dahl, Christian M., Torben S. D. Johansen, Emil N. Sørensen, Christian E. Westermann, and Simon Wittrock, "Applications of machine learning in tabular document digitisation," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 2023, *56* (1), 34–48.

Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in "2009 IEEE conference on computer vision and pattern recognition" IEEE 2009, pp. 248–255.

Fischer, Martin, Martin Karlsson, Therese Nilsson, and Nina Schwarz, "The long-term effects of long terms–compulsory schooling reforms in sweden," *Journal of the European Economic Association*, 2020, *18* (6), 2776–2823.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep Residual Learning for Image Recognition," in "Proceedings of the IEEE conference on computer vision and pattern recognition" 2016, pp. 770–778.

Sutskever, Ilya, James Martens, George Dahl, and Geoffrey Hinton, "On the importance of initialization and momentum in deep learning," in "International conference on machine learning" 2013, pp. 1139–1147.