# Early-life Adversity and Long-term Outcomes: Evidence from the Cocoa Swollen Shoot Virus Outbreak in the Gold Coast

Idris Kambala *

**Abstract**

A large body of work highlights that early-life conditions shape adult outcomes. However, most of this evidence comes from developed economies where safety nets are robust. Moreover, much of the existing literature focuses on broad, economy-wide shocks, with limited attention to sector-specific disruptions. This study addresses these gaps by investigating the long-term effects of early-life exposure to an agricultural epidemic in a low-income setting with weak social protection. Specifically, I examine the long-term impact of the cocoa swollen shoot virus disease (CSSVD) epidemic, which devastated the Gold Coast's (now Ghana) cocoa sector in the 1940s. The CSSVD caused severe yield and income losses and led to compulsory cocoa farm destruction in an environment with little institutional support. Using historical archives and georeferenced survey data, I exploit variation in exposure to the epidemic based on birth timing and ecological suitability for cocoa cultivation to estimate its long-term effects on human capital and labor market outcomes. I find that early-life exposure to the epidemic is associated with persistent human capital deficits. Affected children experienced a 1.1-year loss in schooling (a 23 percent reduction) and a 9.3 percentage point (pp) reduction in attaining any level of formal education (17 percent reduction). They also exhibited a 5.8 pp higher likelihood of morbidity (an 18.8 percent increase) and a 4.6 pp higher probability of living with a disability in adulthood (a 25.7 percent increase). The epidemic, however, had no robust, consistent long-term effects on labor market outcomes.

**Keywords:** early-life exposure, cocoa swollen shoot virus, human capital, agricultural shocks, Ghana

**JEL Classification:** I15, I25, J13, N57, O12

*Department of Economics, University of South Carolina, iddrisu.kambala@grad.moore.sc.edu. www.iddrisukambala.com. All errors and omissions are mine.

## I. Introduction

Why do some individuals thrive while others struggle, even when they grow up in the same country or community? According to the Fetal Origins Hypothesis (FOH), part of the answer may lie in the conditions they faced during early childhood or even before they were born. The FOH, formalized by Barker (1990), posits that environmental conditions and exposures during the prenatal period can cause permanent changes in an individual's health, cognitive development, education, and economic wellbeing.[1] The premise is that fetal development is a critical window during which vital organs form, making them particularly sensitive to adverse shocks. Exposures during this period can therefore lead to permanent physiological or developmental changes, with lasting consequences for outcomes in adulthood (Almond and Currie, 2011). This framework has reshaped how researchers think about inequality because it emphasizes early-life conditions as key determinants of adult wellbeing.

Economists have provided compelling evidence in support of this hypothesis. Studies have linked in-utero and early-childhood exposures to events such as the 1918 Spanish Flu, the Chinese Famine, and the Dutch Hunger Winter to persistent reductions in schooling, height, income, and overall health (Almond, 2006; Almond et al., 2007; Brandt et al., 2016; Chen and Zhou, 2007; Lin and Liu, 2014; Ramirez and Haas, 2022). These findings reinforce the idea that adverse shocks early in life can have long-lasting effects.

Despite these advances, most of the evidence comes from developed countries. There is limited empirical research on fetal origins in developing world contexts, even though these regions are equally, if not more, exposed to diverse shocks that affect child development (Akresh et al., 2012; Maccini and Yang, 2009). Importantly, developed economies typically benefit from stronger social protection systems, which can mitigate the long-term consequences of early-life shocks. In contrast, many developing countries lack such institutional support, potentially exacerbating the effects of these shocks (Beegle et al., 2006; Dercon, 2002).

Moreover, most FOH research has focused on broad shocks such as famines, wars, or natural disasters that have wide-reaching effects, or on health-related shocks that directly compromise human biology (Akbulut-Yuksel et al., 2025; Chen and Zhou, 2007; Currie and Almond, 2011; Lindeboom et al., 2010). Far less is known about industry-specific disruptions, such as agricultural epidemics, which do not directly cause biological harm but instead impose severe economic hardship on households (Banerjee et al., 2010; Hornbeck, 2012). This study addresses these gaps by combining archival records with georeferenced household survey data to investigate the long-term impact of the Cocoa Swollen Shoot Virus Disease (CSSVD) outbreak, a viral agricultural epidemic that devastated cocoa farms in the Gold Coast (now Ghana) during the 1940s. By exploring this context and channel, the study

[1] In some contexts, this is also known as the Developmental Origins Hypothesis.

offers new perspectives on the long-term consequences of early-life adversity.

The CSSVD is a highly destructive, mealybug-transmitted virus that infects cocoa trees at every stage of development. It causes yield losses of up to 70 percent and kills infected trees within 2–3 years (Ameyaw et al., 2015; Lot et al., 1991; Muller, 2008). First reported in 1936 and confirmed as a viral disease in 1939, the outbreak spread rapidly, destroying over half of cocoa trees in some areas by the mid-1940s (Danquah, 2003). Containment required the removal of both infected and nearby trees, leading to the destruction of over 44 million trees between 1946 and 1955 (Thresh and Owusu, 1986). Consequently, farmers faced not only the shock of crop failure but also further losses from compulsory tree removals.

In this study, I investigate how early-life exposure to the CSSVD outbreak affected long-term human capital and labor market outcomes. The shock may have influenced child development through several channels. First, income losses from reduced cocoa production could have limited access to food and other essential resources, resulting in nutritional deficits that impair cognitive development, physical growth, and health (see Almond and Mazumder, 2011). Second, financial strain may have elevated maternal stress during pregnancy, a known risk factor for adverse birth outcomes such as low birth weight and preterm delivery (see Aizer et al., 2016; Case et al., 2005). Third, income constraints could have reduced parental investments in education, healthcare, and early-life stimulation, potentially resulting in poorer child health and delayed educational attainment (see Currie and Hyson, 1999; Royer, 2009). Finally, in response to falling cocoa income, parents may have increased reliance on child labor, thereby limiting children's time for schooling and rest. Against this backdrop, I hypothesize that children exposed to the CSSVD epidemic in utero or during early childhood suffered persistent deficits in education, health, and labor market participation compared to unexposed cohorts.

To identify the long-term effects of CSSVD exposure, I first reconstruct a map of climatically suitable cocoa areas using historical archives. Second, I use GPS coordinates from the 1993 Ghana Demographic and Health Survey (DHS) data to identify respondents residing in cocoa-suitable areas. Third, I define my target cohort as individuals born during the epidemic years or exposed to the crisis during early childhood. I then implement a cohort-based difference-in-differences strategy, where exposure is defined by the interaction between cohort indicators and cocoa-area residency. This strategy leverages variation in exposure across both birth timing and space. I also add year of birth fixed effects in the specification to account for differences across cohorts, and region fixed effects to capture variation across administrative units.

The main findings show that the CSSVD outbreak had a negative impact on human capital, measured by education and health. Children exposed in utero or early childhood attained less schooling and experienced poorer health outcomes. Specifically, exposure is associated with a 1.15-year reduction in schooling – approximately a 23 percent decline relative to the sample mean – and a 9 percentage point (pp) decline in access to formal education, representing about a 17 percent reduction. Morbidity, proxied by recent illness or injury, increased by about 5.8 pp, corresponding to an 18.8 percent rise, and exposure raised the probability of living with a disability in adulthood by 4.6 pp, roughly a 25.7 percent increase. However, the epidemic had no meaningful effect on labor market outcomes. Exposure in utero or early childhood did not significantly alter employment type or labor supply.

The null effect on labor market outcomes is puzzling, given theoretical expectations that education influences labor market performance. One interpretation is that while early-life shocks may persistently affect education and health, which are largely shaped early in life, their effects on more adaptable outcomes such as employment may attenuate over time. In other words, despite early disadvantages, individuals may have eventually integrated into the labor market through alternative pathways such as informal work, on-the-job training, accumulated experience, or new economic opportunities.

I examine whether the results vary by gender and by place of residence. The adverse effects on education and health are concentrated among female children, suggesting that parents may have prioritized investments in boys during the crisis. For rural-urban differences, the epidemic had negative impacts on education in both settings. However, while it increased morbidity and reduced work hours in rural areas, it had no such effects in urban areas. This suggests that pre-existing rural health challenges compounded the epidemic's adverse effects.

In addition to heterogeneity, I test the sensitivity of the results to several robustness checks. The findings are robust to alternative definitions of exposure, including wider and narrower cohort bands. The robustness checks also reveal that timing of exposure matters in that the adverse effects are stronger for individuals exposed in utero or in very early childhood compared to those exposed slightly later. The results also hold when restricting the sample to communities more directly tied to the cocoa economy. I also conduct a falsification test using post-epidemic cohorts as the "exposed" group, finding no effects.

Finally, I assess whether selective migration biases the results. Using respondents' migration histories, I compare movers and non-movers across cocoa and non-cocoa areas. Migration between these regions was modest overall and even lower among the target cohorts. To confirm robustness, I conduct two complementary robustness checks. The first one involves a regression analysis that includes migration indicators and their interactions with cohort exposure, and the second restricts the sample to non-movers. Across both strategies, the baseline results remain robust, suggesting that migration does not meaningfully confound the estimated effects of the CSSVD epidemic.

This study makes three contributions. First, it is the only study to investigate the long-term impacts of the CSSVD outbreak in the Gold Coast. Unlike other cocoa diseases such as capsids or black pods, CSSVD was uniquely destructive, capable of annihilating the entire cocoa industry if left unchecked. Moreover, the timing of the outbreak compounded its severity. It struck just as cocoa had become the country's most valuable export, accounting for approximately 80 % of total exports in 1939 (Gyamera et al., 2023). Furthermore, because cocoa cultivation was dominated by smallholder farmers, the epidemic delivered a highly localized economic shock. This shock likely had profound implications for household welfare and child development, making it a critical episode for understanding how early-life adversity can shape long-term outcomes.

Second, to my knowledge, this is also the first study to examine the long-term developmental effects of an agriculture-related viral epidemic that directly affected cash crop farmers. A closely related study is Banerjee et al. (2010), who analyze the long-lasting impacts of the phylloxera epidemic that destroyed 40 % of French vineyards. They find effects on adult height but not on health or longevity. My study differs in both context and mechanism. Unlike an insect pest, CSSVD was a viral disease that not only destroyed farms directly but also required large-scale destruction of additional farms as part of its control strategy. The dual shock of yield loss and forced tree removal intensified household hardship.

Third, this study contributes to the broader fetal origins of health and human capital (FOH) literature by examining the long-term impacts of an agriculture-based shock, a context that has received comparatively little attention. While most FOH studies focus on health crises, natural disasters, famine, or pollution, this study shows that industry-specific disruptions, such as agricultural disease outbreaks, can also produce lasting effects on education and health. A distinguishing feature of the CSSVD outbreak is its spatial concentration, which contrasts with the diffuse nature of most disasters studied in FOH research. This localized pattern not only allows for more credible identification of causal effects but also enables a clearer understanding of the mechanisms – income loss, food insecurity, unemployment, reduced parental investment, and maternal stress – through which early-life exposure translate into developmental harm. By doing so, it broadens the FOH framework to include economic shocks that operate indirectly through household and market conditions.

The rest of the study proceeds as follows. Section II. provides a historical overview of cocoa production and the onset of the CSSVD outbreak in the Gold Coast. Section III. discusses the empirical strategy and data. Section IV. reports the main empirical estimates, as well as the heterogeneity and robustness checks. Section V. concludes.

## II. Cocoa and the Swollen Shoot Virus in the Gold Coast

Cocoa production in the Gold Coast began in the late 19th century, after a local farmer introduced cocoa seeds acquired during his travels abroad. Initially, cocoa farming was not widely adopted. Cocoa began to gain traction in the mid-to-late 1880s, and it was during this period that the Gold Coast started exporting small quantities. The first recorded export, about 121 lbs, occurred in 1885. There were no further exports until 1891, when another, smaller shipment of 80 lbs was made. The industry's spectacular growth came at the turn of the 20th century, with exports surging from 600 tons in 1900 to becoming the Gold Coast's most valuable export by 1910. By 1911, the Gold Coast had established itself as the world's leading cocoa producer.

Climatic and ecological conditions make the forest zone of the Gold Coast the only suitable area for cocoa cultivation. These cocoa-suitable areas, shown in figure 1 and digitized from (Dickson, 1969), cut across six administrative regions, including Ashanti, Brong Ahafo, Central, Eastern, Western, and Volta. Only a small portion of Greater Accra is suitable, though the region has been historically tied to the cocoa economy through transport infrastructure and export channels. The three northern regions – Northern, Upper East, and Upper West – have no cocoa-suitable areas and are disconnected from the industry.

By the late 19th century, farmers in these cocoa-suitable areas had rapidly adopted cocoa cultivation because it was highly profitable. Cocoa farming was comparatively easier to establish and less costly to maintain than other cash crops such as coffee or oil palm. Starting a cocoa farm required only fertile land and intensive labor at the early stages. Once trees reached fruit-bearing age, they could yield for more than four decades (Gyamera et al., 2023). Consequently, cocoa farming quickly became the primary economic activity in these areas.

Yet the greatest threat to the cocoa boom soon emerged. By the mid-1930s, farmers began to observe escalating dieback on cocoa trees. No one knew what was causing it until 1936, when a farmer brought a sample of an infected cocoa tree – characterized by distended branches and dieback – to the Department of Agriculture to draw attention to the issue. The pathologist at the Department, W.F. Steven, observed that the symptoms were likely caused by an infectious disease, an observation later confirmed by field experiments. The disease would attack every part of the tree – stem, roots, leaves and pods – irrespective of its stage of growth. Once infected, the tree steadily sheds its leaves, loses the characteristic spherical shape of its pods, and eventually dies. Because its most visible symptom was a swollen stem and shoot, it became known as the "swollen shoot" virus disease.

Because the disease was entirely new, its control proved difficult. Initially, Dade, a mycologist, mistakenly attributed the disease to insufficient shading. Based on his advice, the government required cocoa to be planted under tall trees, and farmers who violated the rule were punished. This practice continued until 1939, when the disease was officially confirmed to be viral in nature.

To curb the spread, authorities introduced "roguing," the careful detection and removal of infected cocoa trees. While more effective, this method was still limited because neighboring trees often carried latent infections. As a result, the strategy was expanded to include both symptomatic and adjacent asymptomatic trees. Specifically, trees within 5, 10, and 15 meters of outbreaks of 1–10, 11–100, and more than 100 infected trees, respectively, were to be removed (Thresh et al., 1988).

Based on this diagnosis, the government embarked on a large-scale and continuous tree-removal campaign. Available official data from the mid-1940s offer a sense of the operation's scale. Figure 2 shows the number of diseased trees cut annually from 1945/46 to 1955/56 using data from (Dickson, 1960). Between 1945 and 1950, removals ranged from 52,000 to 1 million trees. After 1951, the figure surged dramatically, peaking at 5.5 million in 1954/55. A more nuanced observation emerges when we examine the per-acre removals. The number of trees cut per acre saw an early spike, reaching a peak of 660 trees/acre in 1947/8. This dropped sharply to 257 trees/acre the following year, and showed only slight increases in subsequent years. Although the overall number of trees continued to rise, the trend from the number of trees cut per acre suggests an initial intensive cutting period, followed by a more controlled or selective cutting approach. Thus, it is likely that tree removals per acre were even higher in the early years of the campaign.

The massive cutting operation provoked strong public discontent. Farmers would prefer to reap smaller harvests from infected farms rather than lose entire farms to compulsory eradication. The loss was even particularly severe for settler farmers, whose landlords could reclaim land at the end of the season (Dzahini-Obiatey et al., 2010). Cocoa farming was also a long-term investment, requiring substantial commitments of land, labor, and credit. And since seedlings needed about seven years to bear fruit, compulsory cutting represented a devastating blow. Nevertheless, in 1946 the government passed legislation making cutting mandatory. Farmers who refused faced arrest and prosecution.[2]

By this time the virus had already spread widely, devastating vast tracts of cocoa farms. Between 1939 and 1944, about two-thirds of the trees planted over the period 1904–1914 were devastated by the virus. Cocoa yield had also plummeted, with output per farm dropping from 30 tons between 1926–1929 to just 6 tons in 1943–1944 seasons. Official statistics suggest that of the 400 million trees in the country as of 1947, 46 million (11.5 %) had been infected and were expected to die within a year. The speed of spread was even more alarming, proceeding at 15 million trees per year. Estimates suggested that, if unchecked, the cocoa industry would disappear within 20 years.[3]

The main objective of my study is to investigate whether early-life exposure to the CSSVD epidemic is associated with long-term deficits in human development and labor market outcomes. The study is grounded in the Fetal Origins Hypothesis (FOH), which posits that adverse conditions during in-utero development and early childhood can have lasting consequences because these are critical periods for physiological and cognitive development.

The epidemic may have affected child development through several interrelated pathways.

[2]The government's hardline stance on tree cutting was informed by expert advice. Worth noting was the recommendation from plant pathologists at the Food and Agriculture Organization (FAO) of the United Nations, who warned that failure to eradicate infected and neighboring trees would lead to the extinction of the cocoa industry. Consequently, compulsory cutting continued.

[3]This discussion draws from Danquah (2003).

The economic loss could have limited access to food and resources needed for nutrition, health, and growth. Reduced family resources may also have constrained parental investments in education, healthcare, and early-life stimulation. Moreover, households may have relied more heavily on child labor, reducing children's time for schooling and rest. Beyond the economic channel, the epidemic may have heightened maternal stress during pregnancy, which is linked to adverse birth outcomes and lifelong deficits.[4] Building on these mechanisms, I hypothesize that early-life exposure to the CSSVD epidemic resulted in persistent adverse effects on education, health, and labor market outcomes.

[4] For further discussion on how parental income and family socioeconomic status affect child development, see Akee et al. (2010); Berger et al. (2009); Blau (1999); Currie (2009); Dahl and Lochner (2012); Duncan et al. (2011); Persson and Rossin-Slater (2018).

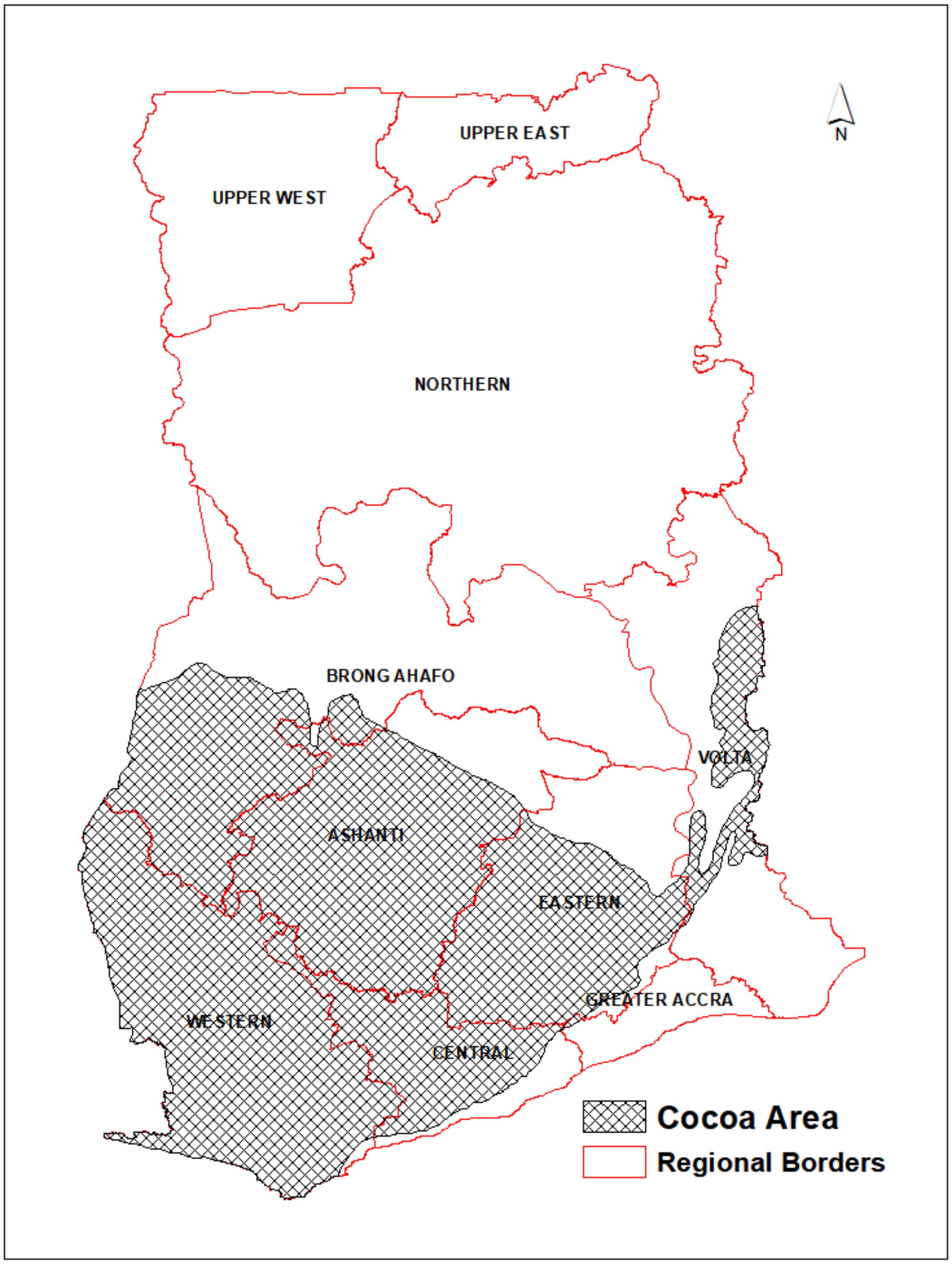


Figure 1: The figure shows the climatically suited cocoa-growing areas and the ten regions of Ghana as of 1993. The gray-shaded cocoa areas are constructed using data from (Dickson, 1969).

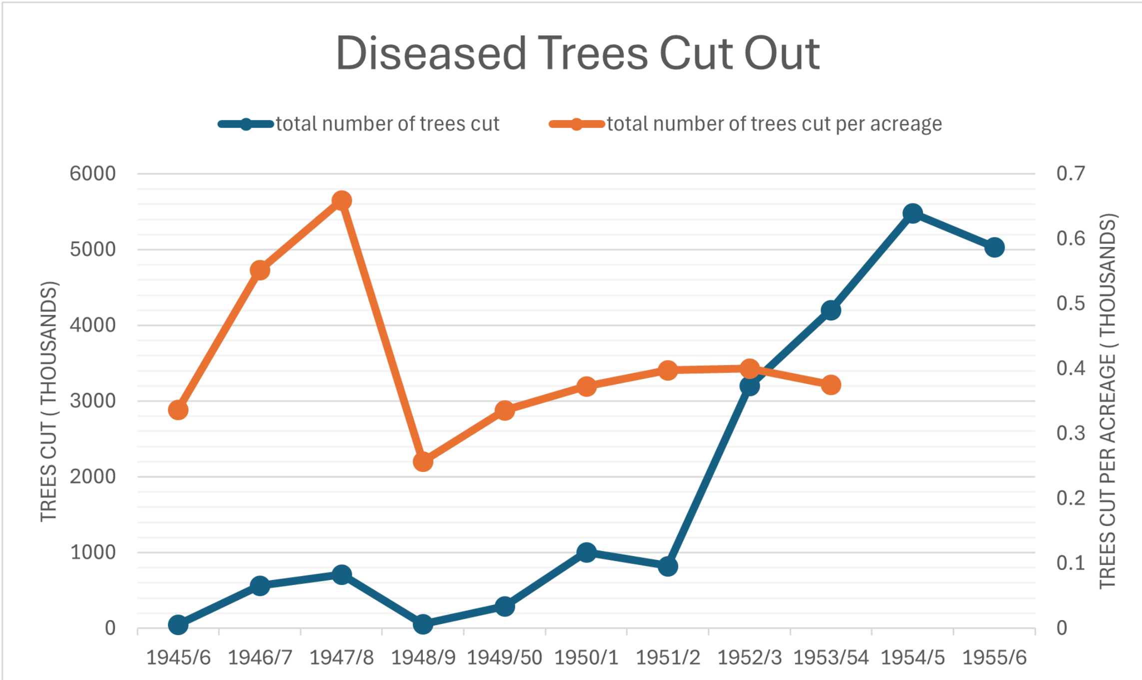


Figure 2: The figure shows the number of diseased trees cut annually from 1945/6 to 1955/6. Data from (Dickson, 1960)

## III. Data and Methodology

In this section I describe the data used for the study, the target and exposed cohorts, and the empirical framework used to identify the long-term effect of early-life exposure to CSSVD.

### III.I Data

The main data I use to investigate the long-term effect of early-life exposure to the CSSVD epidemic is the 1993 Ghana Demographic and Health Survey (DHS).[5] The DHS is a household-level, nationally representative survey conducted primarily in low- and middle-income countries. With support from USAID, the survey is conducted periodically by national governments and gathers comprehensive information on health indicators including fertility, family planning, maternal and child health, mortality rates, nutrition, and HIV/AIDS. The DHS also includes information on other socioeconomic indicators such as domestic violence, education, and labor market variables.

DHS survey data are organized into recodes in order to simplify analyses. For this study, I use the Household Member Recode (HHMR), focusing on human capital and labor market variables. I proxy human capital by education and health, measuring the former by years

[5] Although the first DHS survey conducted in Ghana was in 1988, this round not only lacks some of the variables of interest, but more importantly, it also lacks GPS and locality information necessary for spatial mapping. The 1993 DHS was the next round of survey.

of schooling (in single years) and having formal education, and the latter by morbidity and disability. Specifically, morbidity indicates whether the respondent reported being ill or injured in the two weeks prior to the interview. Disability indicates whether the respondent was living with at least one of several forms of disabilities. These include whether the respondent was experiencing difficulty in moving, seeing, hearing, speaking, learning, or has loss feeling in hand or foot, or suffered from fits, or behaved strangely, or faced any other form of difficulty.

For the labor market variables, I look at two broad outcomes. The first is employment type, which indicates whether the respondent was either self-employed, employer, unpaid family worker, or waged employed (i.e. worked for someone else for wages or salary). I focus on waged and self-employment. The second variable is labor supply. For this, I focus on the maximum number of hours per day worked in the past seven days, and the number of days the respondent worked out of the last seven days.

### III.II Identifying the Target and Exposed Cohorts

To accurately identify the long-term effects of early-life exposure to the CSSVD epidemic, I identify two groups, namely the target and exposed cohorts. The target cohort consists of individuals who were born or were under five years of age during the height of the epidemic, while the exposed cohort is the subset of this group who were living in cocoa-suitable areas during those years.

The timing of the epidemic provides a clear guide for identifying these groups. As stated earlier, the viral nature of the epidemic was confirmed in 1939, and by the mid-1940s the disease had spread extensively, causing severe devastation to cocoa farms and substantial income losses for cocoa-dependent households. This period of intense impact was further compounded by wartime neglect, as World War II diverted government attention and resources away from addressing the crisis. Thus, the 1940s represent the most critical period of widespread cocoa yield collapse and household income loss.

Based on this context, I define the target cohort as individuals born between 1936 and 1949. This range captures those who were either exposed in utero or during the critical early childhood years (under age five) at the time of the epidemic's most intense phase. Given the target cohort, the exposed cohort is the subset of individuals within this birth window who were located in cocoa-suitable areas.

The rationale for selecting this birth window (1936–1949) requires further econometric, institutional, and economic elaboration. From an econometric perspective, selecting a narrower window may introduce bias if it excludes children who were plausibly exposed to the epidemic. Conversely, a wider window that includes "distant" or marginally affected cohorts could dilute the treatment effects. The 1936–1949 window strikes a balance because it includes those most likely to have experienced the direct consequences of the epidemic

while minimizing contamination from cohorts whose exposure was minimal or mitigated.
The institutional justification relates to the timing of monetary compensation for affected farmers. During the early years of the epidemic, cocoa households bore the full cost of forced tree removal without any reimbursement. Even after limited compensation was introduced in the late 1940s, payments were minimal and covered only the removal of seemingly healthy neighboring trees, not the infected ones. It was not until June 1951 that the government announced the "New Deal for Cocoa," which offered the most substantial and generous compensatory payments to date. Thus, individuals born in the 1950s were unlikely to have been exposed to the full cost of the epidemic. In the robustness section, I show that individuals born in the 1950s were indeed not affected.
From an economic perspective, cocoa households would have likely adapted gradually to the shock over time. As the epidemic persisted, farmers would have sought alternative income sources or coping mechanisms, such as crop diversification, informal labor, or migration, in order to mitigate the impact of the CSSVD outbreak. These coping mechanisms, coupled with government support in the later part of the epidemic, would have helped households partially mitigate the impact of the crisis. Therefore, focusing on the 1936–1949 birth cohorts ensures that I capture individuals who were exposed before significant adaptation or policy intervention could alter the severity of the crisis' developmental effects.

### III.III Empirical Framework

My primary objective is to estimate the long-term impact of early-life exposure to the CSSVD epidemic on human capital and labor market outcomes. To identify the effect, I employ a cohort-based Difference-in-Differences (DID) design that exploits variation across both birth cohorts and spatial exposure to the CSSVD outbreak:

$$\begin{aligned} y_{ibr} = & \sum_{b=1928}^{1968} \gamma_b \, \text{BirthYear}_b + \alpha \, \text{CocoaArea}_i \\ & + \beta \left( \text{TargetCohort}_i \times \text{CocoaArea}_i \right) + \mathbf{FE}_r + \varepsilon_{ibr} \end{aligned} \tag{1}$$

where $y_{ibr}$ denotes the long-term outcome for individual $i$, born in birth year $b$ and residing in administrative region $r$. $\sum_{b=1928}^{1968} \gamma_b \text{BirthYear}_b$ represents a full set of year-of-birth dummies for $b \in 1928, ..., 1968$. These cohorts would have been between 25–65 years old at the time of the 1993 DHS survey. I focus on this age band because it ensures that all respondents had completed their formal education and had reached an age where long-term health and labor market outcomes can be reasonably assessed.
These year-of-birth fixed effects ensure that comparisons are made within cohorts, not across them. For example, educational attainment may have improved over time due to nationwide investments in schooling. The year-of-birth fixed effects absorb such cohort-specific

common shocks, allowing me to isolate the specific impact of being exposed to the CSSVD epidemic from broader trends.

The term $\text{CocoaArea}_i$ is a binary indicator equal to 1 if individual $i$ resides in an area that is climatically suitable for cocoa production and 0 otherwise. As depicted in Figure 1, I reconstruct cocoa-suitable zones using a historical map from Dickson (1960). To classify respondents as living within or outside these areas, I use the GPS coordinates of the DHS sampling units. This classification is central to the identification strategy, since only individuals in cocoa-suitable areas were plausibly exposed to the economic shocks caused by the CSSVD epidemic.

As the cocoa–noncocoa map (Figure 1) illustrates, the boundaries of cocoa-suitable areas do not align with Ghana's ten administrative regions.[6] Therefore, I include a full set of dummies for the ten administrative regions, $\mathbf{FE}_r$, to absorb time-invariant region-specific factors that could affect child development. For example, some regions might have systematically better access to schools or health facilities, independent of the CSSVD epidemic.

Including both the cocoa-area indicator and region fixed effects strengthens identification. This setup allows comparisons across individuals born in the same region and same birth cohort but with differing exposure to the CSSVD epidemic. At the same time, it leverages variation between cocoa-growing and non-growing regions. Thus, this structure not only controls for regional confounding but also exploits fine-grained variation in spatial exposure.

The interaction term ($\text{TargetCohort}_i \times \text{CocoaArea}_i$) identifies my exposed group, the key cohort of interest. $\text{TargetCohort}_i$ is an indicator equal to 1 if individual $i$ was born between 1936 and 1949. By interacting the target cohort with the $\text{CocoaArea}_i$ indicator, I identify individuals exposed to the CSSVD epidemic in utero or during early childhood. The coefficient $\beta$ captures the additional long-term impact of being exposed to the epidemic in cocoa-producing areas, relative to individuals born outside the target window and/or outside cocoa-suitable areas.

It is important to emphasize that the CSSVD crisis imposed two layers of disruption. First, the viral infection itself damaged cocoa trees, reducing yields and threatening the livelihoods of farmers. Second, the government's compulsory tree-removal campaign, which extended beyond infected trees to include adjacent healthy ones, further intensified the economic shock by forcing mass destruction of productive assets. Specification 1, however, does not separate these channels. Instead, it captures the bundled effect of both the disease and the state's containment response. Likewise, while various mechanisms may underlie the observed adult outcomes the empirical model does not disentangle their individual contributions. As such, the estimates should be interpreted as the combined long-term impact of early-life exposure to the CSSVD crisis.

Lastly, one possible concern with this empirical setup is migration. Individuals may have

[6]The map reflects Ghana's ten regions as of 1993. Subsequent political demarcations may vary.

moved across or between cocoa and non-cocoa areas after birth, which could blur the true measure of exposure. I address this issue explicitly in the robustness checks section, where I show that migration between cocoa and non-cocoa areas was relatively limited and does not materially bias the estimated effects.

### III.III.I Descriptive Statistics

The descriptive statistics of the respondents in my sample are presented in Table 1. Columns (1) to (5) display the statistics for the whole sample, while columns (6) to (7) and columns (8) to (9) report the statistics for the target cohort and all other cohorts, respectively. The total number of observations ranges from 6,214 to 7,070. The missing values are concentrated in the labor market outcomes, which is likely due to recall problems or the exclusion of individuals, such as retirees and children, who are not in the labor force.

The average age of respondents in the sample is 40 years. Educational attainment is relatively low, with respondents averaging about 5 years of schooling. However, the majority (55 percent) have had some form of formal education. Roughly three out of ten people reported recent morbidity, while 18 percent reported living with some form of disability. A large share of the sample (61 percent) are household heads, and the average age of household heads is about 44 years. Although more than half of the sample is female (56 percent), only 26 percent of household heads are women. The data also show that nearly half of respondents (46 percent) reside in cocoa-growing areas, and rural residents account for a disproportionately high share (67 percent) of the sample. Lastly, the target cohort makes up about 26 percent of the sample.

Some noticeable disparities exist between the target cohorts and other cohorts for certain outcomes. The target cohorts have fewer years of schooling, are less likely to have received formal education, report higher disability, and are more likely to experience recent morbidity. By contrast, there are no remarkable disparities in labor market outcomes. My empirical strategy therefore interacts indicators of the target cohorts and cocoa areas to determine the differential impact of CSSVD exposure on human capital and labor market outcomes.

Table 1: Summary Statistics

| | (1) N | (2) min | (3) max | (4) mean | (5) sd | (6) N | (7) mean | (8) N | (9) mean |
|---|---|---|---|---|---|---|---|---|---|
| | Whole Sample | | | | | Target Cohorts | | Other Cohorts | |
| Years of Schooling | 7,067 | 0 | 20 | 4.986 | 5.161 | 1,814 | 3.687 | 5,253 | 5.434 |
| Have Formal Education | 7,070 | 0 | 1 | 0.550 | 0.498 | 1,815 | 0.404 | 5,255 | 0.600 |
| Morbidity | 7,070 | 0 | 1 | 0.309 | 0.462 | 1,815 | 0.349 | 5,255 | 0.295 |
| Disability | 7,070 | 0 | 1 | 0.179 | 0.383 | 1,815 | 0.254 | 5,255 | 0.153 |
| Days Worked | 6,214 | 0 | 7 | 5.024 | 1.780 | 1,641 | 5.071 | 4,573 | 5.007 |
| Hours Worked | 6,220 | 0 | 24 | 7.148 | 2.894 | 1,643 | 7.116 | 4,577 | 7.160 |
| Wage Employment | 6,219 | 0 | 1 | 0.170 | 0.375 | 1,642 | 0.166 | 4,577 | 0.171 |
| Self Employment | 6,219 | 0 | 1 | 0.715 | 0.451 | 1,642 | 0.735 | 4,577 | 0.708 |
| Age | 7,070 | 25 | 65 | 40.09 | 11.42 | 1,815 | 50.09 | 5,255 | 36.64 |
| Year of Birth | 7,070 | 1928 | 1968 | 1953 | 11.42 | 1,815 | 1943 | 5,255 | 1956 |
| Cocoa Area | 7,070 | 0 | 1 | 0.462 | 0.499 | 1,815 | 0.456 | 5,255 | 0.464 |
| Female | 7,070 | 0 | 1 | 0.564 | 0.496 | 1,815 | 0.576 | 5,255 | 0.560 |
| Head of Household | 7,070 | 0 | 1 | 0.611 | 0.487 | 1,815 | 0.667 | 5,255 | 0.592 |
| Age of Head of Household | 7,069 | 17 | 95 | 44.22 | 12.92 | 1,815 | 52.80 | 5,254 | 41.26 |
| Female Head | 7,070 | 0 | 1 | 0.260 | 0.439 | 1,815 | 0.275 | 5,255 | 0.255 |
| Rural | 7,070 | 0 | 1 | 0.671 | 0.470 | 1,815 | 0.699 | 5,255 | 0.662 |

The table displays the descriptive statistics of respondents in the sample. The cocoa indicator refers to places climatically suited for cocoa cultivation. The target cohorts are all individuals born between 1940 and 1950.

## IV. Results

In this section I report and discuss the empirical estimates of early-life exposure to the CSSVD outbreak. I start with the baseline estimates, followed by heterogeneous analysis, and robustness checks. To allow for arbitrary correlation of errors within both cohorts and regions, I cluster the standard errors at the cohort × region level.

### IV.I Baseline Estimates

The baseline estimates of the long-term impacts of CSSVD exposure on human capital and labor market outcomes are reported in Tables 2 and 3, respectively. All models include year-of-birth and region fixed effects. The odd-numbered columns contain no controls, whereas the even-numbered columns include controls for age of household head and indicators for female, household head, female head, and rural residence. I focus on the conditional estimates in the even-numbered columns.

The results in Table 2 demonstrate that early-life exposure to the CSSVD outbreak had a significant and persistent negative impact on human capital outcomes. Although cocoa-suitable areas generally exhibit better educational outcomes overall, exposure to the epidemic is associated with worse outcomes. The estimates in column (2) show that early-life exposure is associated with a 1.15-year loss in schooling, corresponding to about a 23 percent reduction relative to average schooling. Similarly, being exposed in utero or during early childhood leads to a 9.3 percentage point (pp) reduction in the probability of having some level of formal education (column 4), equivalent to about a 17 percent drop from the average.

The results also show that early-life exposure to CSSVD is associated with long-term health deficits. The estimates in columns (6) and (8) indicate that exposed children were about 5.8 pp and 4.6 pp more likely to report recent morbidity and some type of disability in adulthood, respectively. These correspond to 18.8 percent and 25.7 percent increase above their respective averages.

Turning to the labor market outcomes in Table 3, the results suggest that early-life exposure to the CSSVD epidemic did not have statistically significant long-term effects on individuals' labor market outcomes. Across all models, the coefficients of interest are small in magnitude and not statistically different from zero. In other words, the epidemic neither affected exposed children's long-term employment type (waged or self employment) nor their labor supply (days or hours worked).[7]

These results patterns are further visually confirmed in Figures 3 to 10. The figures plot residualized outcomes by birth cohort, separately for individuals in cocoa and non-cocoa areas. I include year-of-birth and region fixed effects as well as all the other controls. The figures for human capital outcomes (figures 3 to 6) show clear divergence for cohorts born during the epidemic years. In cocoa areas, we observe notable declines in schooling and formal education, and increases in morbidity and disability, although the disability pattern is somewhat noisy. By contrast, the labor market outcomes (figures 7 to 10) reveal no clear, consistent divergence between cocoa and non-cocoa cohorts for employment type or labor supply.

Taken together, the results suggest that while early-life exposure to the CSSVD had lasting negative effects on human capital, it did not affect labor market outcomes. These findings highlight one important implication. It suggests that while early-life shocks may have persistent effects on education and health, this does not necessarily translate to disadvantages in the labor market. One possible explanation is that outcomes such as education and health are determined during critical developmental stages, and thus early shocks leave permanent marks. By contrast, the labor market provides greater flexibility and opportunities for adaptation through mechanisms such as informal employment, job training, skill acquisition, and work experience. These channels may compensate for early disadvantages over time. In other words, although education is generally a strong determinant of labor market success, the adaptability of labor markets allows individuals to overcome the disadvantages imposed by early-life shocks.

[7] I also rerun the regressions using natural logs of work hours and days, finding similar results.

### IV.II Heterogeneous Effects

The effects of early-life exposure to the CSSVD epidemic may not have affected all exposed individuals uniformly. To assess whether the effects varied across key demographic groups, I split the sample by gender and by type of place of residence (rural vs. urban). Examining these subgroups separately helps uncover whether certain populations were disproportionately affected by the epidemic.[8]

#### IV.II.I Effects By Gender

The gender-specific results are presented in figure 11. They reveal notable heterogeneity in the long-term effects of CSSVD exposure. Generally, the effects are primarily driven by female children. For educational outcomes, the effect for males is not statistically significant, whereas the effect for females is negative and significant. Similarly, for health outcomes, the adverse effects of CSSVD exposure are concentrated among females. Most labor market outcomes are not statistically significant for either gender, with the exception of wage employment for females, which suggest that the epidemic reduced long-term female labor market participation.

Overall, the results indicate that the long-term effects of the CSSVD epidemic were not uniform across gender. Exposed female children experienced significantly worse educational outcomes than their male peers, and these educational deficits appear to have limited their access to wage employment in adulthood. These findings imply that in the face of economic hardship during the epidemic, households may have prioritized investments in boys over girls.

#### IV.II.II Effects By Type of Place of Residence

Figure 12 shows whether the long-term effects of CSSVD exposure differ by place of residence, specifically between individuals in rural and urban areas. For educational outcomes, CSSVD exposure is associated with substantial reductions in years of schooling and the likelihood of formal education in both rural and urban areas. Health outcomes show more distinct rural–urban differences. The adverse health effects of CSSVD exposure are concentrated among rural residents. Exposed rural children experienced significant increases in both morbidity and disability, with significantly coefficients than the baseline estimates. There are no significant differences observed among exposed urban children. For labor market outcomes, most coefficients are insignificant, except for work hours among rural

[8] I do not conduct heterogeneity analysis for other control variables such as female-headed households or household head status because they are not strictly time-invariant and may not reflect the individual's early-life context. For example, currently residing in a female-headed household does not necessarily imply that the individual was raised in such a household. While rural–urban status is also time-varying, it serves as a reasonable proxy for one's early-life context.

residents, indicating that early-life exposure to CSSVD slightly reduced work hours.

In sum, while rural–urban differences did not substantially shape the impact of CSSVD exposure on educational outcomes, they did matter for health and labor supply. Rural residents bore the brunt of the epidemic's effects on health and work hours, experiencing higher morbidity and disability as well as reduced labor supply. These findings point to compound vulnerability, in that the pre-existing limited health infrastructure and poor sanitation conditions in rural areas amplified the negative effects of the epidemic on health outcomes.

### IV.III Robustness Checks

I conduct a series of sensitivity checks to test the robustness of my results. These include adjusting the comparison cohort bands, specifying alternative definitions of treatment cohorts, restricting the analysis to cocoa-affiliated regions, conducting a falsification test using post-epidemic cohorts, and applying multiple strategies to account for potential bias from selective migration.

#### IV.III.I Alternative Cohort Bands

My first robustness check examines the sensitivity of the estimates to the choice of comparison cohorts. I re-estimate the baseline results using narrower and wider bands of birth years. Instead of the original window of 1928–1968, I restrict the sample first to individuals born between 1930–1960 and then to those born between 1920–1970. This allows me to test whether the baseline results are driven by cutoff choices around the target cohort.

The main results, reported in Panels A and B of Appendix Table A1, remain unchanged. Early-life exposure to the CSSVD epidemic continues to have significant and persistent impacts on educational and health outcomes. These findings suggest that the observed effects are not an artifact of the specific cohort range used in the baseline analysis.

#### IV.III.II Disaggregating the Treated Cohort by Exposure Timing

Next, I assess whether the long-term effects of the CSSVD epidemic differ by timing of early-life exposure. I disaggregate the exposed cohorts into two subgroups: "only early childhood" and "in utero to early childhood." The first group includes individuals who were 0–5 years old (exclusive of age five) during the epidemic's peak, while the second group includes those exposed from in utero through early childhood. This analysis allows me to assess whether developmental timing shapes the intensity of the effects.

The results, reported in Appendix Table A2, show that the adverse educational effects persist and are similar for both groups. In both cases, exposure reduced years of schooling by about 1.2 years and lowered the probability of receiving formal education by 9 pp. For health outcomes, the epidemic increased the incidence of morbidity for those exposed from in-utero through early childhood. The effects on disability and labor market outcomes are not statistically significant. The only exception is its effect on work hours for the in-utero to early childhood group. For this group, the epidemic marginally decreased work hours per week by 0.28 hours.

Overall, the epidemic affected both groups with similar magnitudes, but prenatal exposure appears to drive the health effects. This finding is consistent with the FOH literature, which emphasizes that the earliest stages of life are the most sensitive periods of vulnerability, particularly for health.

#### IV.III.III Comparison Within Cocoa-Affiliated Regions

In the baseline analysis, the comparison group included individuals born either outside cocoa areas or in cocoa-suitable areas but outside the target cohort. In this robustness check, I restrict the sample by excluding the three northern regions – Upper West, Upper East, and Northern – which have no historical, ecological, or economic ties to the cocoa sector. This restriction strengthens identification by ensuring that comparisons are made only within regions closely tied to the cocoa economy.

Excluding the northern regions also addresses a historical confounder. colonial policies restricted the expansion of western education in the north, creating significant educational disparities with the south. These disparities persisted even after independence (Kambala, 2023). Removing the northern regions from the sample therefore helps isolate the long-term effects of CSSVD exposure from broader north–south historical development differences. The resulting sample includes seven regions, all within the southern ecological belt.[9]

The results, reported in Appendix Table A3, remain broadly consistent with the baseline. Early-life exposure to CSSVD continues to negatively affect schooling, access to formal education, and morbidity. Specifically, exposed individuals completed 0.87 fewer years of schooling, were 5.8 pp less likely to have formal education, and were 7 pp more likely to report recent morbidity relative to unexposed individuals in cocoa-affiliated regions. The coefficients for disability and labor market outcomes are small and not statistically significant. In sum, the findings suggest that the observed effects are not driven by broader north–south disparities or unrelated contextual differences.

#### IV.III.IV Falsification Test Using Post-Epidemic Birth Cohorts

I have argued that because the height of the epidemic occurred between 1940 and 1949, only children exposed early childhood or born during this period would have been most exposed. As a falsification exercise, I examine later-born cohorts, specifically 1950-–1960,

[9] Although only a small portion of the Greater Accra region is cocoa-suitable, I retain it in the sample due to its historical role as a logistical and administrative hub for cocoa transport and export.

who were unlikely to be directly affected by the epidemic. This allows me to test whether the estimated effects persist for individuals with minimal or no plausible exposure.

Although the CSSVD remained prevalent even in the 1950s, there are two main reasons why its impact would have diminished by this period. As previously discussed, first, by this time, cocoa farmers would have adopted diverse coping strategies to mitigate losses. Second, again as discussed before, beginning in the early 1950s the government implemented more generous compensatory payments to cocoa farmers, including the "New Deal for Cocoa" announced in 1951. Therefore, individuals born between 1950 and 1960 were largely insulated from the full economic burden of the epidemic.

The results of this falsification exercise, reported in Appendix Table A4, confirm this expectation. The 1950–1960 birth cohorts were not systematically affected by the epidemic. Most point estimates are small and not statistically different from zero. The only exception is morbidity, which is statistically significant but has a counterintuitive sign suggesting that CSSVD reduced morbidity. This may reflect a post-epidemic rebound, where recovery efforts (such as compensation) and coping strategies improved health outcomes. Overall, these findings confirm that the effects documented in this study reflect exposure to the epidemic itself, not underlying trends in cocoa areas or particular cohorts.

### IV.III.V Threat to Identification: Selective Migration

In this exercise I examine whether selective migration might have biased the main estimates. Since I classify individuals as exposed or unexposed based on their current place of residence, migration could lead to misclassification of exposure and introduce upward or downward bias. For example, estimates could be downward biased if out-migrants were negatively selected and/or stayers were positively selected. They could also be downward biased if the effects of the epidemic persisted for exposed children who subsequently migrated to non-cocoa areas, since their worse outcomes would pull down the average outcomes of the truly unexposed group.

I first assess the prevalence of selective migration in my sample by exploring respondents' migration histories. DHS survey asks respondents to provide information about their region of birth if they were not "currently" residing there. Using this, I am able to "repatriate" respondents to their regions of birth and examine the share of migrants and non-migrants in cocoa and non-cocoa regions as well as their demographic characteristics. Note that in this exercise I can only identify whether a respondent's region of birth is a cocoa-producing region or not. I cannot identify their exact birth locality to confirm if it is cocoa-suitable. This implies that I am only capturing long-distance migration, which typically involves major life decisions and, more importantly, is the most likely source of bias. It must also be emphasized that while this question is useful for assessing the demography of migrants and non-migrants, it cannot be used to gauge their socioeconomic status during the epidemic years, since the survey was conducted in 1993.

Panels A and B of Appendix Table A5 present the summary statistics of migration patterns. The table shows that overall migration rates were high. About 70.7 percent of respondents reported moving from a different region to their current region of residence. However, the relevant type of migration for this study, which is movement between cocoa and non-cocoa regions, was relatively low. Only 11.2 percent of respondents moved from cocoa to non-cocoa regions, and an even lower share (8.9 percent) moved from non-cocoa to cocoa regions. More importantly, the most consequential migration type for this study involves the target cohorts, since they determine the direction of bias. Panel B shows that members of the target cohorts who moved between cocoa and non-cocoa regions account for only 5.2 percent of the full sample (369 out of 7,070). Considering only the target group, 20.3 percent of them relocated between cocoa and non-cocoa regions – 11.2 percent from cocoa regions to non-cocoa regions and 9.2 percent from non-cocoa regions to cocoa regions. This pattern suggests that migration among the target cohort was modest overall, making it unlikely to significantly bias the main estimates.

To understand if there is systematic selection between migrants and non-migrants among the target cohorts in both cocoa and non-cocoa regions, I simple linear regressions exploring the demographics of the two groups. The results are presented in Panels C and D of Appendix Table A5. Panel C suggests that exposed individuals who moved from cocoa to non-cocoa regions were less likely than their non-migrant counterparts to be female, to reside in rural areas in the destination, or to live in female-headed households there. However, the two groups do not differ significantly in terms of their own age, the age of their household heads, or household headship status. For the target cohort that moved from non-cocoa to cocoa regions (Panel D), the estimates indicate they were less likely than non-migrants to reside in rural areas in their destinations or to live in households with younger household heads there, but more likely to be household heads themselves or to live in female-headed households after moving.

These patterns suggest only modest selection, since core traits such as age and gender are largely similar across groups. Most of the observed differences relate to destination characteristics, which are already controlled for in the main estimates. Any potential bias from migration is therefore also likely modest.

Next, I estimate a regression model that explicitly accounts for selective migration. In this model, I add separate indicators for out-migrants from cocoa and non-cocoa regions and interact them with the target birth cohorts. This model not only accounts for the potential effects of selective migration but also allows the coefficients of the interaction terms to reveal how outcomes differed for target cohorts who migrated relative to those who did not. Results are shown in Appendix Table A6.

The coefficient estimates for the indicator terms suggest that those who moved from cocoa to non-cocoa regions, compared to stayers in cocoa regions, had higher morbidity and were less likely to be waged employed (more likely self-employed). Those who moved from non-cocoa to cocoa regions had worse educational outcomes and were less likely to be self-employed compared to their non-migrant counterparts.

While these patterns suggest some degree of selective migration, they are not informative about whether such selection affected the treated cohort and, hence, the baseline estimates. The interaction terms provide insights in this regard. The interaction terms suggest that for most outcomes, exposed children who moved from cocoa to non-cocoa regions did not differ significantly from their non-migrant peers. However, they reported higher disability rates and longer working hours in adulthood. By contrast, target cohorts who moved from non-cocoa to cocoa regions experienced some gains in educational attainment relative to non-migrants.

These results deserve further discussion. The higher disability rates among exposed individuals who moved from cocoa to non-cocoa regions suggest that households likely migrated after developmental damage had already occurred, so migration did not mitigate the full effects of early-life exposure. This negative selection of out-migrants from cocoa areas partly explains the smaller effects of the epidemic on health outcomes in the baseline results.

The longer working hours observed for out-migrants from cocoa regions also suggest that migration may have triggered behavioral adaptations. Families might have moved to non-cocoa regions in search of better labor opportunities, and longer hours may reflect coping mechanisms to recover from the economic shock of the epidemic. If these adaptations were passed intergenerationally, they could explain the increased labor effort observed in adulthood.

The finding that educational attainment improved for target cohorts moving from non-cocoa to cocoa regions is not surprising. As pointed out earlier, the non-cocoa regions (the three northern regions) historically had limited access to education due to colonial policies. Thus, northern children who migrated south into cocoa regions likely did so for better educational opportunities, which translated into improved attainment compared to their non-migrant peers.

Turning back to the main impact of the epidemic in this new specification, the adverse long-term effects of CSSVD exposure on education and health remain robust. Moreover, the magnitude of the estimates is similar to the baseline. In sum, although modest selective migration is evident, it does not significantly bias the main estimates.

Finally, I further test robustness by re-estimating the main specification on a restricted sample excluding all individuals who migrated between cocoa and non-cocoa regions. This ensures that exposure status is not confounded by later geographic mobility. Appendix Table A7 shows that the main results remain robust. Early-life exposure to CSSVD continues to have statistically significant and economically meaningful long-term negative effects on education and health. These findings confirm that the main effects reflect the lasting consequences of early-life exposure to the epidemic rather than migration patterns.

Table 2: The Long-term Effects of CSSVD Exposure on Human Capital

| | (1) Years of Schooling | (2) Years of Schooling | (3) Have formal Education | (4) Have formal Education | (5) Morbidity | (6) Morbidity | (7) Disability | (8) Disability |
|---|---|---|---|---|---|---|---|---|
| Cocoa Area | 1.157*** | 1.528*** | 0.124*** | 0.149*** | -0.003 | -0.008 | -0.003 | -0.005 |
| | (0.226) | (0.187) | (0.019) | (0.017) | (0.017) | (0.017) | (0.014) | (0.014) |
| Target Cohort × Cocoa Area | -1.088*** | -1.145*** | -0.089*** | -0.093*** | 0.058** | 0.058** | 0.048** | 0.046** |
| | (0.271) | (0.239) | (0.024) | (0.022) | (0.025) | (0.025) | (0.023) | (0.023) |
| Observations | 7,067 | 7,066 | 7,070 | 7,069 | 7,070 | 7,069 | 7,070 | 7,069 |
| R-squared | 0.249 | 0.363 | 0.287 | 0.355 | 0.029 | 0.039 | 0.102 | 0.105 |
| Cohort F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Region F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | No | Yes | No | Yes | Yes | Yes | Yes | Yes |
| Outcome Mean | 4.986 | 4.986 | 0.550 | 0.550 | 0.309 | 0.309 | 0.179 | 0.179 |

The Long-term Effects of CSSVD Exposure on Human Capital. Cocoa area indicates respondents residing in cocoa-suitable areas. Target cohort refers to individuals born between 1935 and 1949. All models include year-of-birth and region fixed effects. The even-numbered columns include controls for age of household head and indicators for female, rural residence, female head of household, and head of household. Standard errors are clustered at the cohort × region level. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: The Long-term Effects of CSSVD Exposure on Labor Market Outcomes

| | (1) Waged employment | (2) Waged employment | (3) Self employment | (4) Self employment | (5) Days worked | (6) Days worked | (7) Hours worked | (8) Hours worked |
|---|---|---|---|---|---|---|---|---|
| Cocoa Area | -0.016 | 0.017 | 0.013 | -0.015 | -0.156* | -0.121 | -0.628*** | -0.459*** |
| | (0.017) | (0.015) | (0.018) | (0.018) | (0.079) | (0.079) | (0.150) | (0.140) |
| Target Cohort × Cocoa Area | 0.006 | 0.002 | -0.011 | -0.008 | 0.073 | 0.069 | -0.162 | -0.188 |
| | (0.021) | (0.018) | (0.025) | (0.023) | (0.101) | (0.100) | (0.163) | (0.153) |
| Observations | 6,219 | 6,218 | 6,219 | 6,218 | 6,214 | 6,214 | 6,220 | 6,219 |
| R-squared | 0.073 | 0.221 | 0.046 | 0.105 | 0.078 | 0.094 | 0.062 | 0.122 |
| Cohort F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Region F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Outcome Mean | 0.170 | 0.170 | 0.715 | 0.715 | 5.024 | 5.024 | 7.148 | 7.148 |

The Long-term Effects of CSSVD Exposure on labor market outcomes. Cocoa area indicates respondents residing in cocoa-suitable areas. Target cohort refers to individuals born between 1935 and 1949. All models include year-of-birth and region fixed effects. The even-numbered columns include controls for age of household head and indicators for female, rural residence, female head of household, and head of household. Standard errors are clustered at the cohort × region level. *** p<0.01, ** p<0.05, * p<0.1.

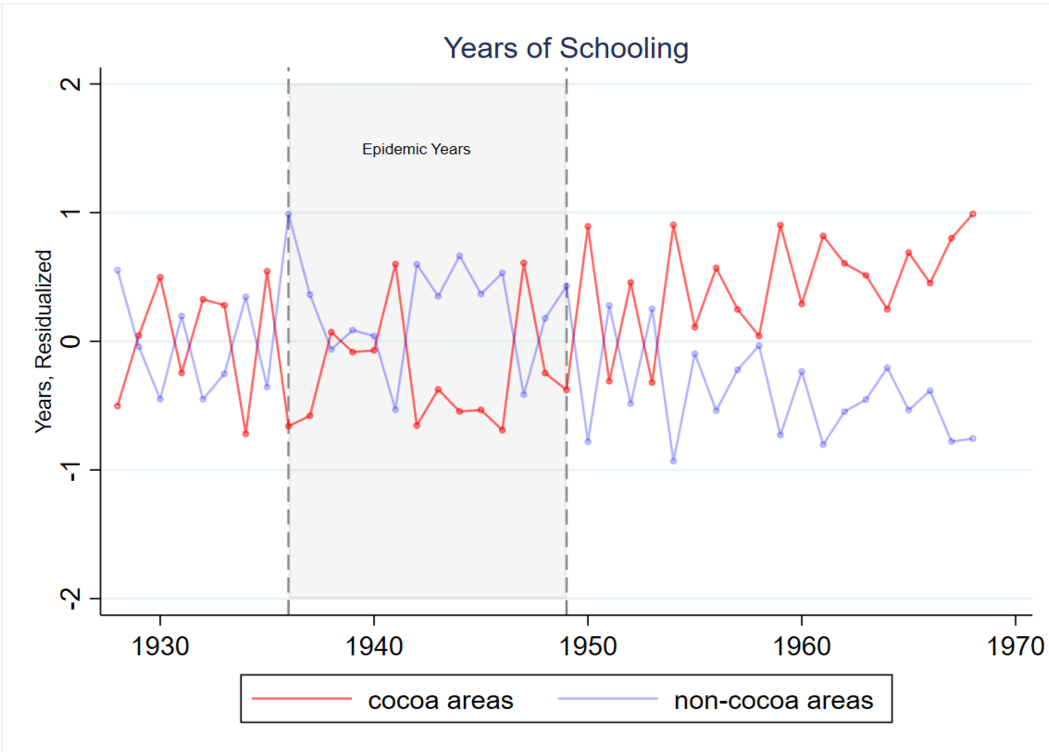


Figure 3: Effect of CSSVD exposure on years of schooling.

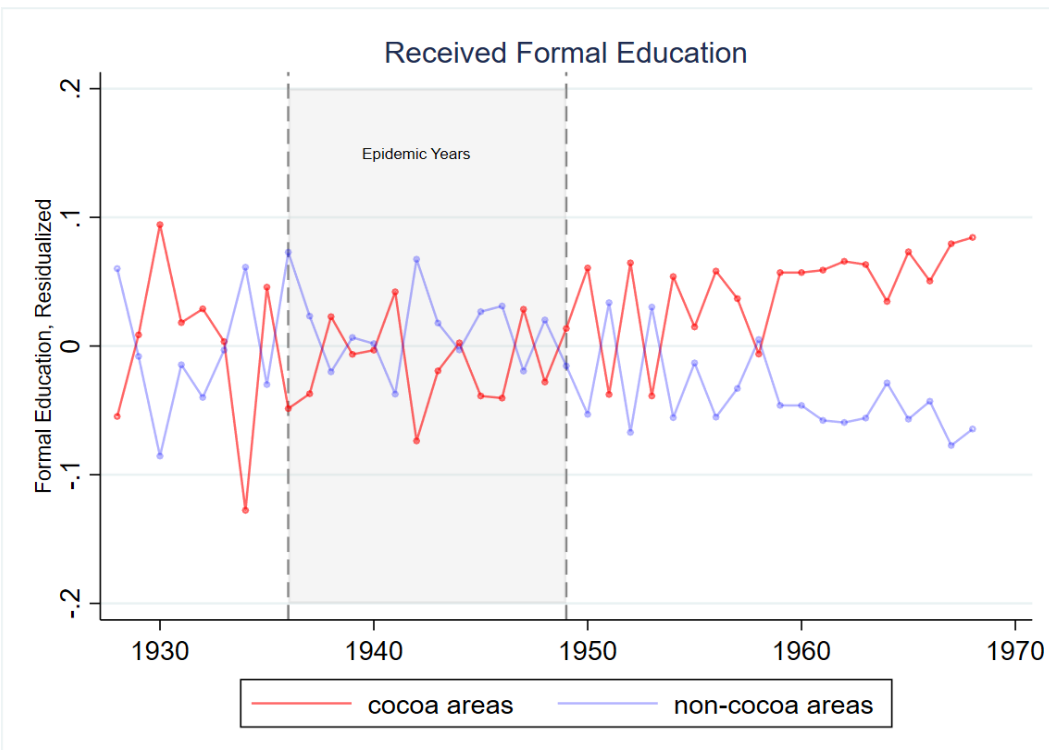


Figure 4: Effect of CSSVD exposure on formal education attainment.

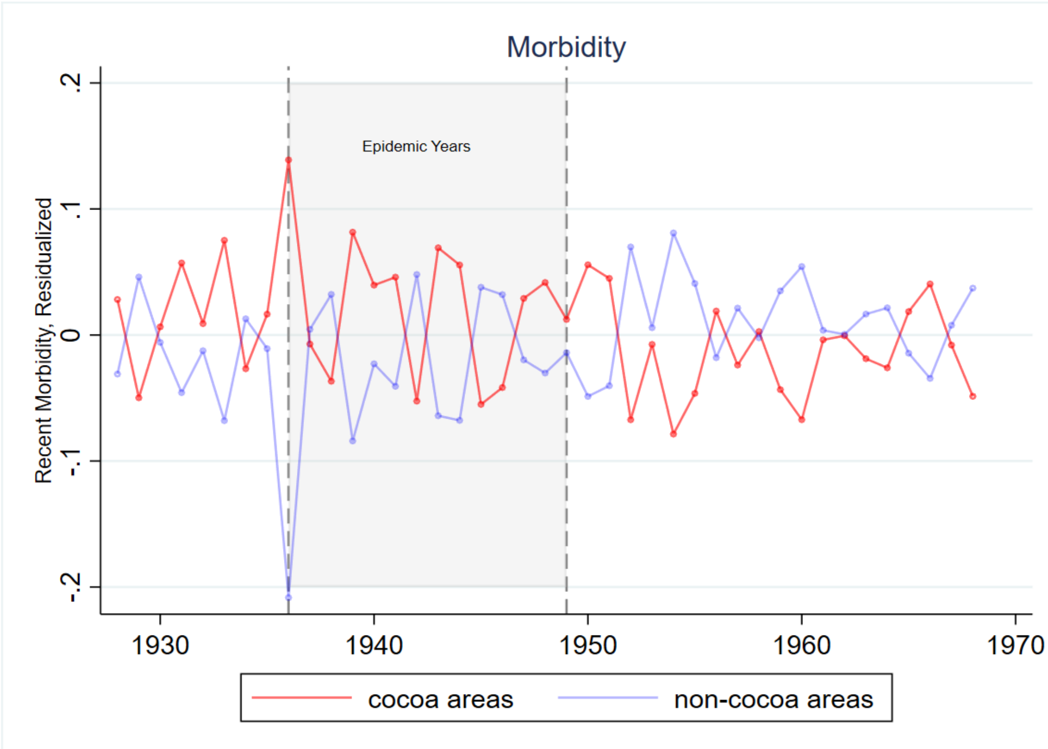


Figure 5: Effect of CSSVD exposure on morbidity in adulthood.

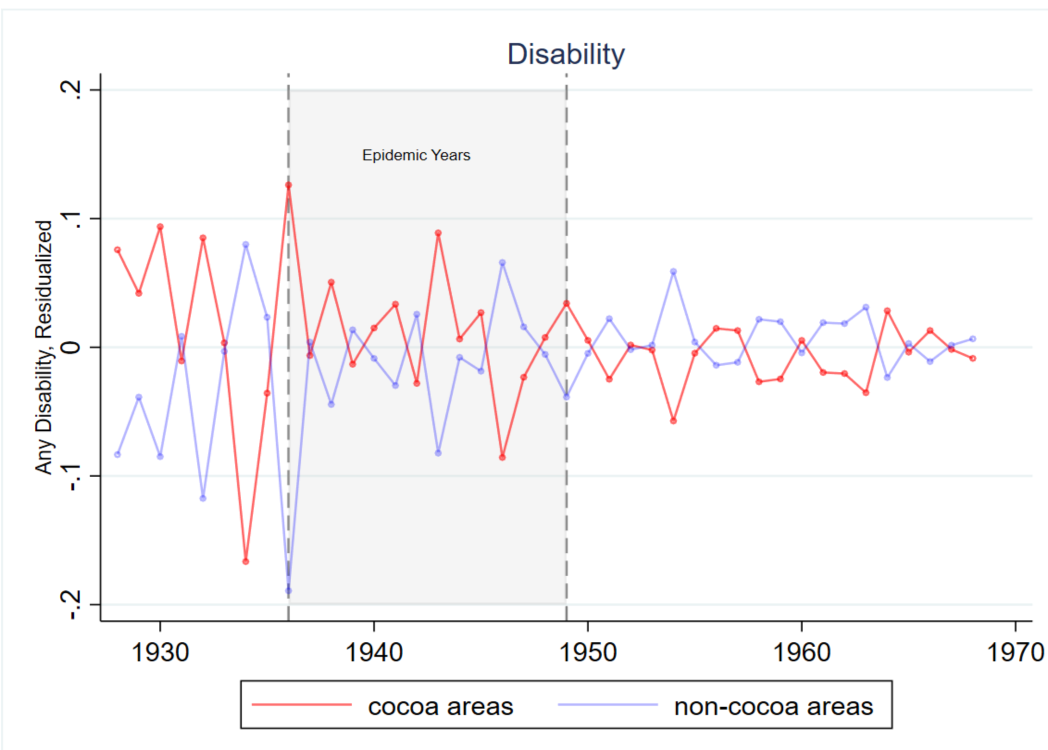


Figure 6: Effect of CSSVD exposure on adult disability.

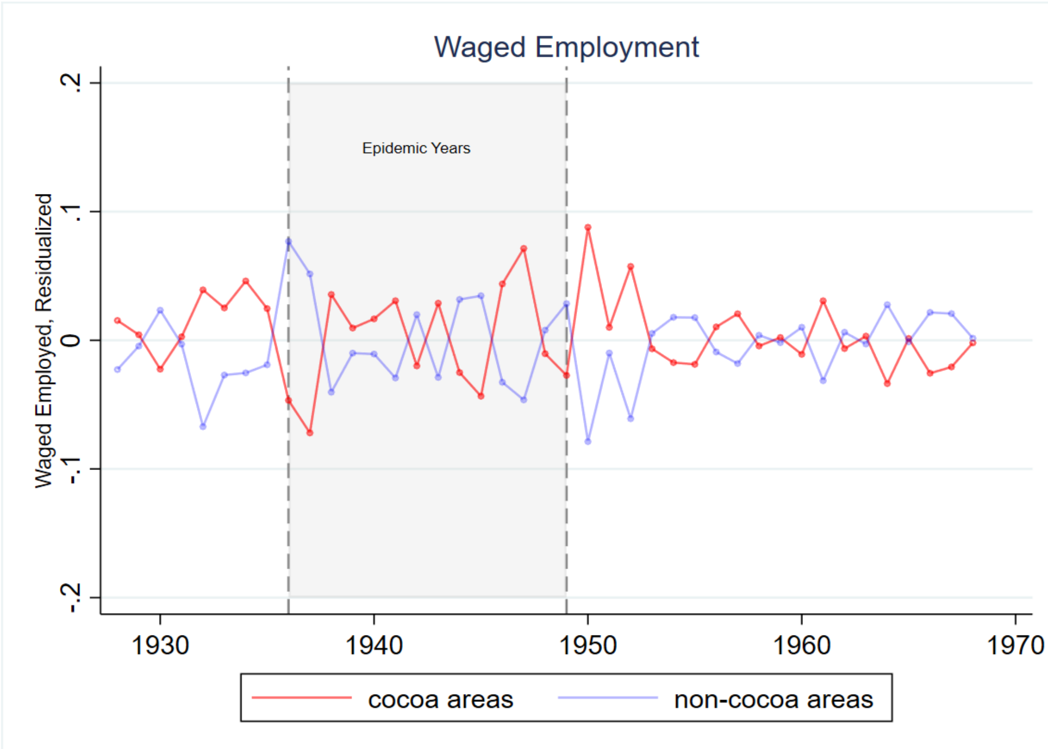


Figure 7: Effect of CSSVD exposure on waged employment.

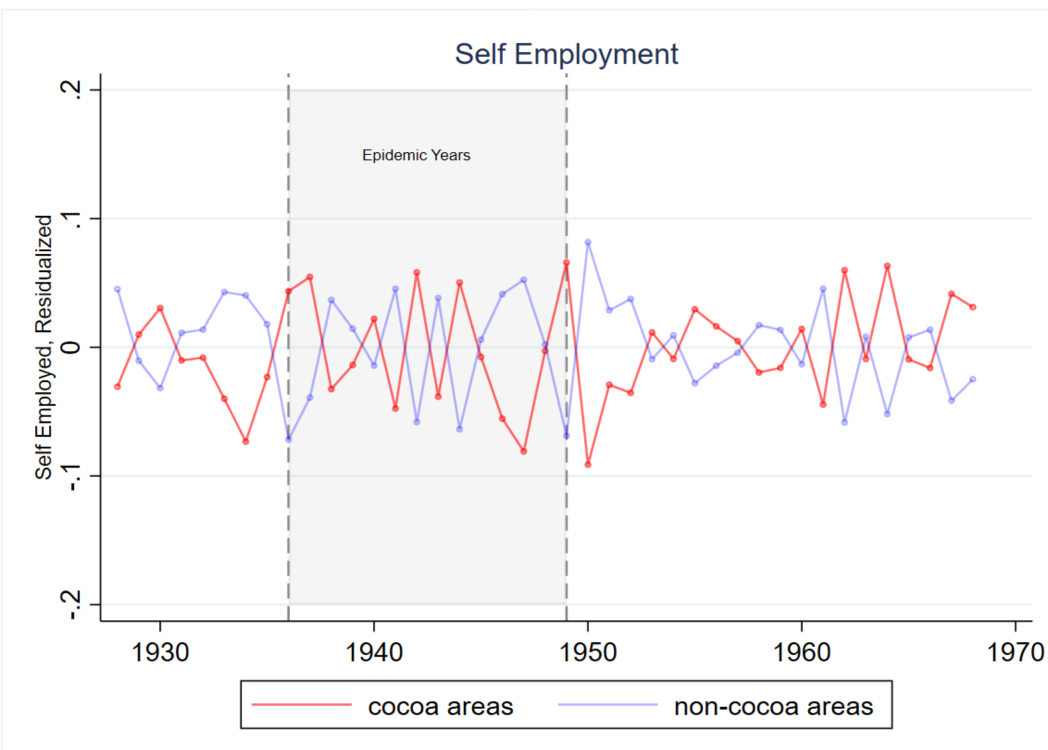


Figure 8: Effect of CSSVD exposure on self-employment.

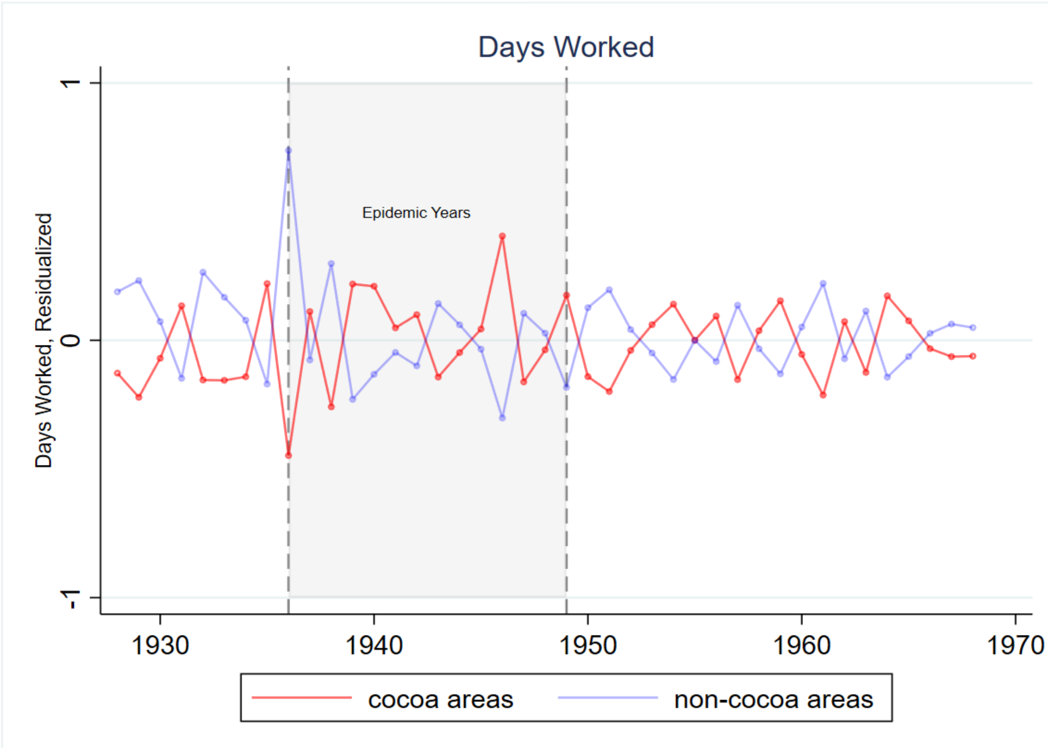


Figure 9: Effect of CSSVD exposure on number of days worked.

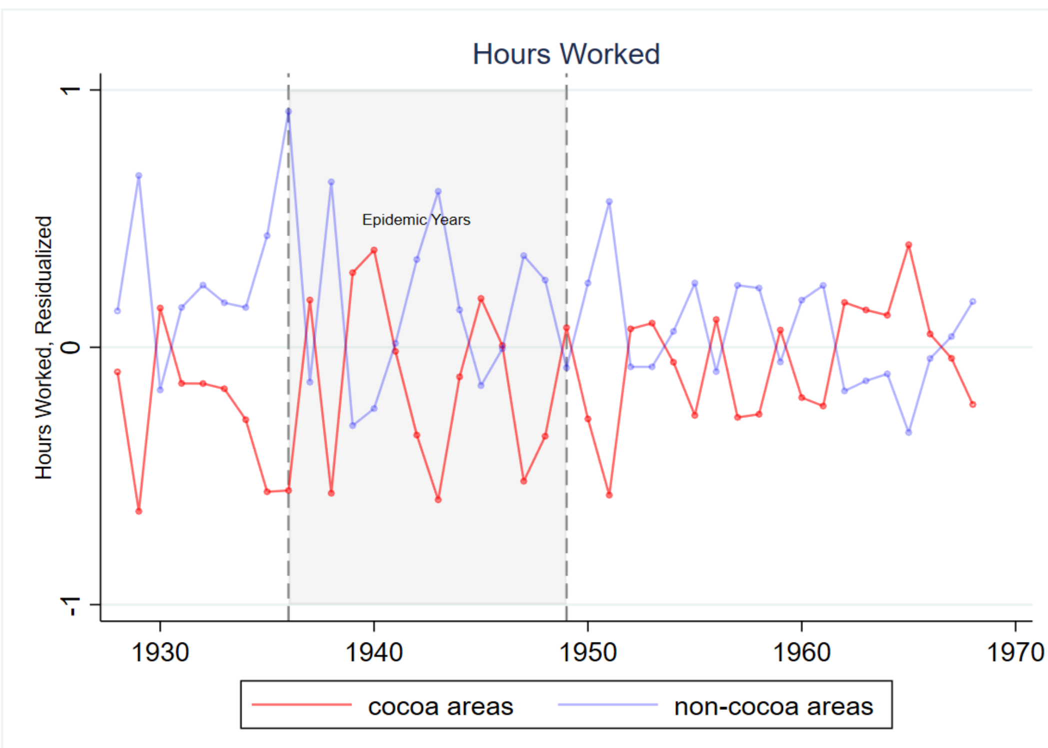


Figure 10: Effect of CSSVD exposure on hours worked per day.

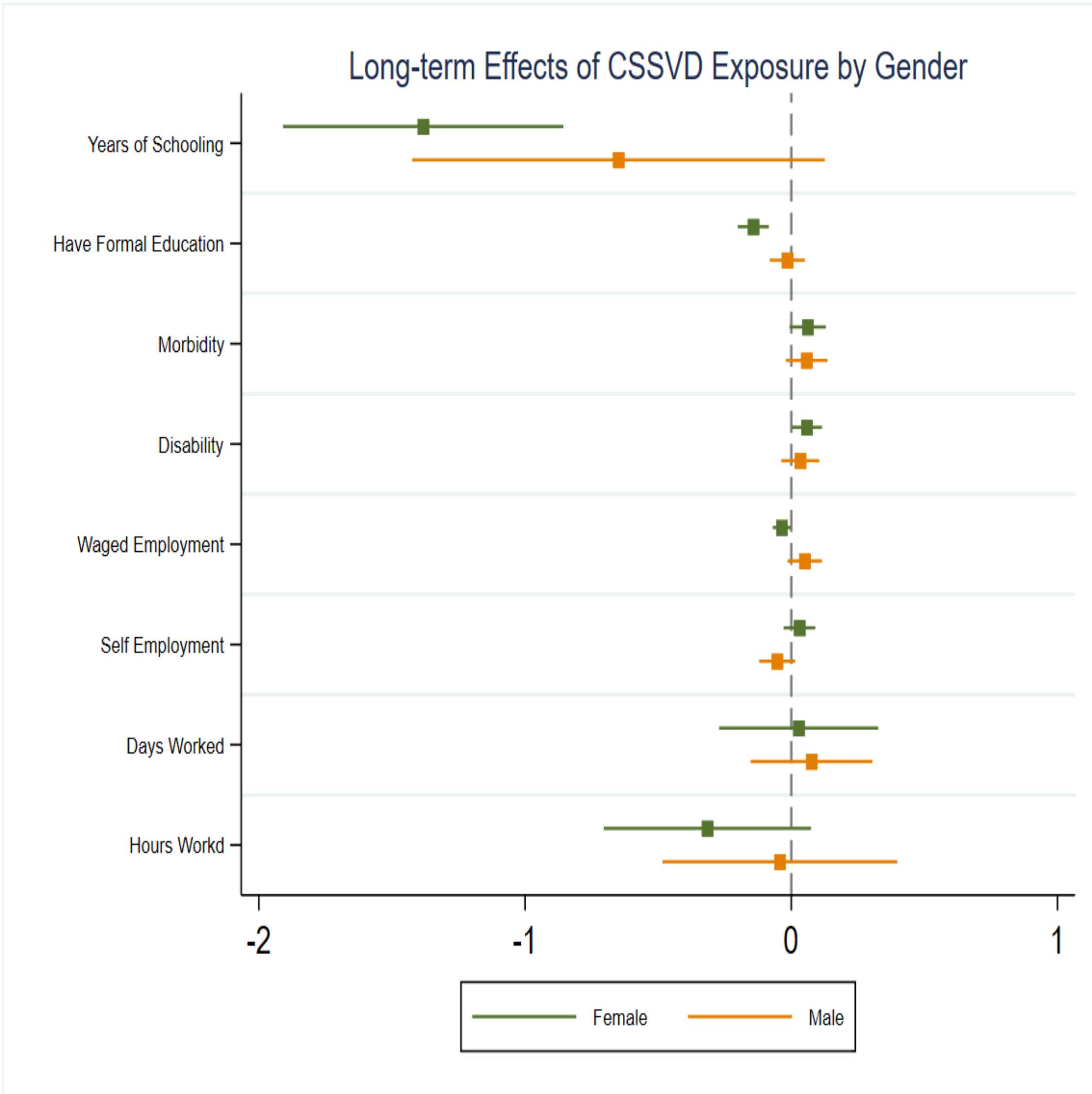


Figure 11: The figure shows the long-term effects of CSSVD exposure by gender. All models include year of birth and region fixed effects as well as controls for age of household head, household head status, gender of household head, and rural-urban residence.

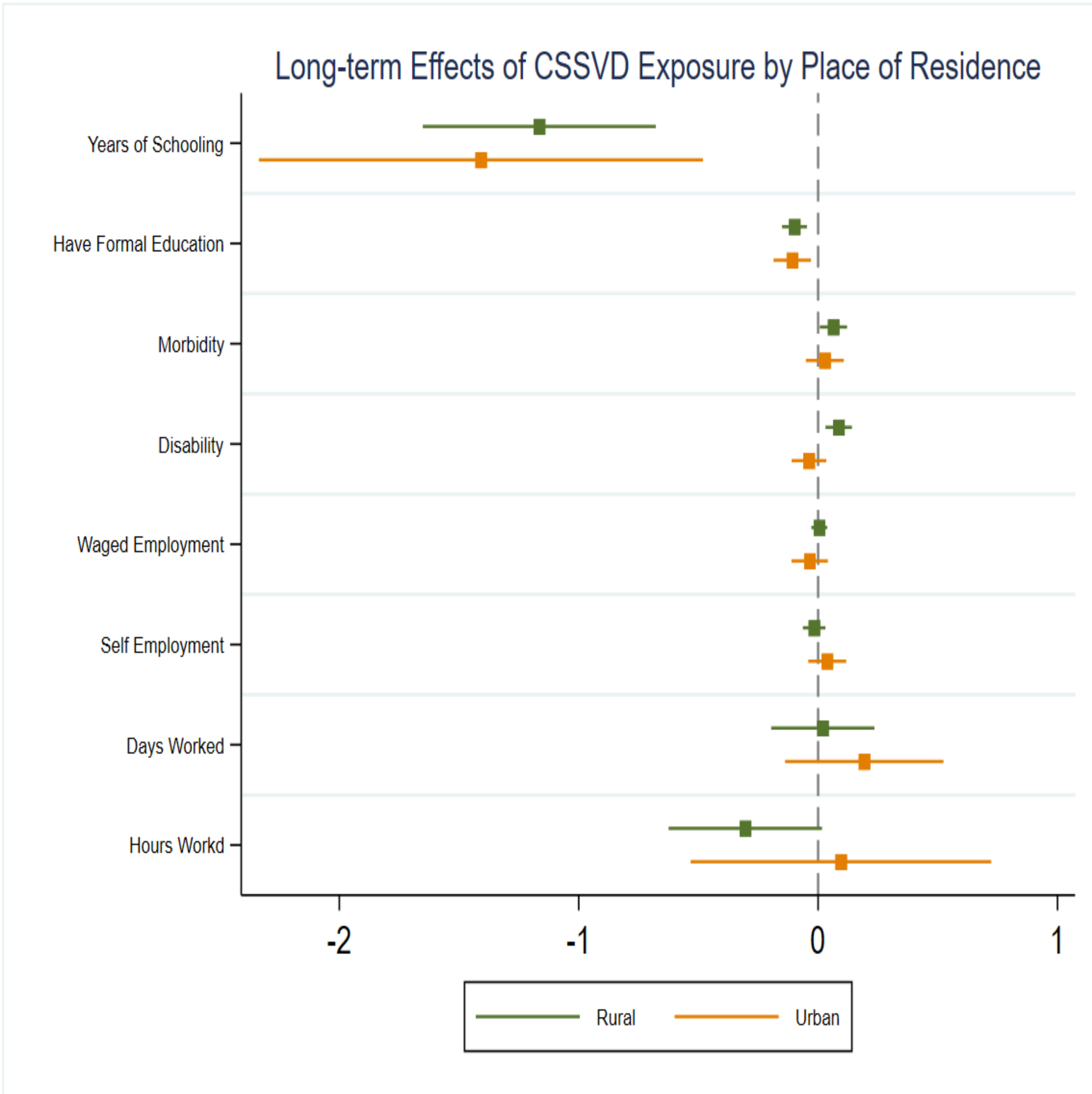


Figure 12: The figure shows the long-term effects of CSSVD exposure by type of place of residence (rural vs. urban). All models include year of birth and region fixed effects as well as controls for gender, age of household head, household head status, and gender of household head.

## V. Conclusion

Cocoa was introduced to the Gold Coast, now Ghana, at the end of the 19th century. By the early 20th century, cocoa farming had become the primary economic activity in the forest zone, where climatic and ecological conditions are suitable for its cultivation. Because of its comparatively low costs of maintenance and high profitability, cocoa farming attracted many farmers who had previously engaged in other crops such as oil palm and coffee.

However, beginning in the 1930s, the cocoa industry faced a major threat as farmers began noticing widespread dieback of cocoa trees caused by a virus. This virus, later termed the cocoa swollen shoot virus disease (CSSVD), causes branches to swell and eventually kills the tree. The disease spread alarmingly fast and caused devastating yield losses. By the mid-1940s, more than 40 million trees had been infected, and the virus continued to spread at an estimated 15 million trees annually. Farm output also plummeted, falling from 30 tons to 6 tons.

After several field experiments, it was determined that controlling the disease required "roguing," the careful detection and removal of infected trees. However, this method proved insufficient, since asymptomatic but infected trees remained. Authorities therefore mandated the removal of both infected trees and their neighbors. As a result, cocoa farmers suffered not only from yield losses due to the virus but also from further economic impacts caused by compulsory tree removals.

In this study, I investigate the long-term effects of this agricultural epidemic on the human capital and labor market outcomes of children exposed in utero or during early childhood. The study is framed within the fetal origins hypothesis, which posits that early-life exposures, such as income shocks and prenatal stress, can have persistent effects on later-life outcomes. In line with this hypothesis, I examine whether the economic shock and its accompanying stressors induced by the CSSVD had lasting consequences for exposed children. To identify the effect, I combine archival records of climatically suitable cocoa-cultivation areas with georeferenced individual-level survey data to identify exposed children based on their birth cohort and residence. I then implement a cohort-based Difference-in-Differences strategy that exploits variation across both birth timing and spatial exposure.

The results show that early-life exposure to CSSVD produced lasting deficits in human capital, proxied by education and health. Exposed individuals lost just over one year of schooling – about a 23 percent decline relative to the average – and were 9 percentage points (pp) less likely to obtain formal education, a 17 percent reduction. They also faced a 5.8 pp higher likelihood of morbidity, roughly an 18.8 percent increase, and a 4.6 pp higher probability of disability in adulthood, corresponding to a 25.7 percent rise. Heterogeneity analysis shows that these adverse impacts were concentrated among female children and rural residents, indicating that the epidemic disproportionately affected society's most vulnerable groups.

However, I find no robust long-term impact of the epidemic on labor market outcomes. Early-life exposure to CSSVD neither affected employment type nor labor supply in adulthood. This null effect on labor market outcomes carries one important implication. It suggests that unlike education and health, which are often strongly shaped by early-life conditions, labor market outcomes allow greater flexibility for adaptation. Despite facing significant early disadvantages in education and health, affected individuals may have eventually integrated into the labor market through avenues such as informal employment, job training, accumulated work experience, or economic opportunities that arose later in life.

## References


Aizer, A., Stroud, L., and Buka, S. (2016). Maternal stress and child outcomes: Evidence from siblings. *Journal of Human Resources*, 51(3):523–555.

Akbulut-Yuksel, M., Tekin, E., and Turan, B. (2025). World war ii blues: The long–lasting mental health effect of childhood trauma. *Journal of Human Resources*. .Published online before print October 8, 2025.

Akee, R. K., Copeland, W. E., Keeler, G., Angold, A., and Costello, E. J. (2010). Parents' incomes and children's outcomes: a quasi-experiment using transfer payments from casino profits. *American Economic Journal: Applied Economics*, 2(1):86–115.

Akresh, R., Lucchetti, L., and Thirumurthy, H. (2012). Wars and child health: Evidence from the eritrean–ethiopian conflict. *Journal of development economics*, 99(2):330–340.

Almond, D. (2006). Is the 1918 influenza pandemic over? long-term effects of in utero influenza exposure in the post-1940 us population. *Journal of political Economy*, 114(4):672–712.

Almond, D. and Currie, J. (2011). Killing me softly: The fetal origins hypothesis. *Journal of economic perspectives*, 25(3):153–172.

Almond, D., Edlund, L., Li, H., and Zhang, J. (2007). Long-term effects of the 1959-1961 china famine: Mainland china and hong kong. Working Paper 13384, National Bureau of Economic Research.

Almond, D. and Mazumder, B. (2011). Health capital and the prenatal environment: the effect of ramadan observance during pregnancy. *American Economic Journal: Applied Economics*, 3(4):56–85.

Ameyaw, G. A., Dzahini-Obiatey, H. K., Domfeh, O., Oppong, F. K., and Abaka-Ewusie, K. (2015). History and data analyses of 'cutting out' method for cocoa swollen shoot virus disease (cssvd) control in ghana. *Journal of Plant Diseases and Protection*, 122:200–206.

Banerjee, A., Duflo, E., Postel-Vinay, G., and Watts, T. (2010). Long-run health impacts of income shocks: Wine and phylloxera in nineteenth-century france. *The Review of Economics and Statistics*, 92(4):714–728.

Barker, D. J. (1990). The fetal and infant origins of adult disease. *BMJ: British Medical Journal*, 301(6761):1111.

Beegle, K., Dehejia, R. H., and Gatti, R. (2006). Child labor and agricultural shocks. *Journal of Development economics*, 81(1):80–96.

Berger, L. M., Paxson, C., and Waldfogel, J. (2009). Income and child development. *Children and youth services review*, 31(9):978–989.

Blau, D. M. (1999). The effect of income on child development. *Review of Economics and Statistics*, 81(2):261–276.

Brandt, L., Siow, A., and Vogel, C. (2016). Large demographic shocks and small changes in the marriage market. *Journal of the European Economic Association*, 14(6):1437–1468.

Case, A., Fertig, A., and Paxson, C. (2005). The lasting impact of childhood health and circumstance. *Journal of health economics*, 24(2):365–389.

Chen, Y. and Zhou, L.-A. (2007). The long-term health and economic consequences of the 1959–1961 famine in china. *Journal of health economics*, 26(4):659–681.

Currie, J. (2009). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *JoUrnal of economIc lIteratUre*, 47(1):87–122.

Currie, J. and Almond, D. (2011). Human capital development before age five. In Card, D. and Ashenfelter, O., editors, *Handbook of Labor Economics*, volume 4, pages 1315–1486.

Elsevier.

Currie, J. and Hyson, R. (1999). Is the impact of health shocks cushioned by socioeconomic status? the case of low birthweight. *American economic review*, 89(2):245–250.

Dahl, G. B. and Lochner, L. (2012). The impact of family income on child achievement: Evidence from the earned income tax credit. *American Economic Review*, 102(5):1927–1956.

Danquah, F. K. (2003). Sustaining a west african cocoa economy: Agricultural science and the swollen shoot contagion in ghana, 1936-1965. *African economic history*, (31):43–74.

Dercon, S. (2002). Income risk, coping strategies, and safety nets. *The World Bank Research Observer*, 17(2):141–166.

Dickson, K. (1960). *Cocoa in Ghana*. Phd thesis, University of London, London.

Dickson, K. (1969). *A Historical Geography of Ghana*. Cambridge University Press, London.

Duncan, G. J., Morris, P. A., and Rodrigues, C. (2011). Does money really matter? estimating impacts of family income on young children's achievement with data from random-assignment experiments. *Developmental psychology*, 47(5):1263.

Dzahini-Obiatey, H., Domfeh, O., and Amoah, F. (2010). Over seventy years of a viral disease of cocoa in ghana: from researchers' perspective. *African Journal of Agricultural Research*, 5(7):476–485.

Gyamera, E. A., Domfeh, O., and Ameyaw, G. A. (2023). Cacao swollen shoot viruses in ghana. *Plant Disease*, 107(5):1261–1278.

Hornbeck, R. (2012). The enduring impact of the american dust bowl: Short-and long-run adjustments to environmental catastrophe. *American Economic Review*, 102(4):1477–1507.

Kambala, M. I. (2023). Colonial origins of comparative development in ghana. *The journal of development studies*, 59(2):188–208.

Lin, M.-J. and Liu, E. M. (2014). Does in utero exposure to illness matter? the 1918 influenza epidemic in taiwan as a natural experiment. *Journal of health economics*, 37:152–163.

Lindeboom, M., Portrait, F., and Van den Berg, G. J. (2010). Long-run effects on longevity of a nutritional shock early in life: the dutch potato famine of 1846–1847. *Journal of health economics*, 29(5):617–629.

Lot, H., Djiekpor, E., and Jacquemond, M. (1991). Characterization of the genome of cacao swollen shoot virus. *Journal of General Virology*, 72(7):1735–1739.

Maccini, S. and Yang, D. (2009). Under the weather: Health, schooling, and economic consequences of early-life rainfall. *American economic review*, 99(3):1006–1026.

Muller, E. (2008). Cacao swollen shoot virus. In Rao, G., S.M.P., K., and Lenardon, S., editors, *Characterization, Diagnosis & Management of Plant Viruses*, volume 1: Industrial Crops, pages 423–444. Studium Press LLC.

Persson, P. and Rossin-Slater, M. (2018). Family ruptures, stress, and the mental health of the next generation. *American economic review*, 108(4-5):1214–1252.

Ramirez, D. and Haas, S. A. (2022). Windows of vulnerability: consequences of exposure timing during the dutch hunger winter. *Population and Development Review*, 48(4):959–989.

Royer, H. (2009). Separated at girth: Us twin estimates of the effects of birth weight. *American Economic Journal: Applied Economics*, 1(1):49–85.

Thresh, J., Owusu, G., and Ollennu, L. (1988). Cocoa swollen shoot: An archetypal crowd disease/afrikanische sproßschwellung an kakaobäumen: Eine archetypische" crowd"-krankheit. *Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz/Journal of Plant Diseases and Protection*, pages 428–446.

Thresh, J. M. and Owusu, G. (1986). The control of cocoa swollen shoot disease in ghana: an evaluation of eradication procedures. *Crop protection*, 5(1):41–52.

## Appendix

## A Tables and Figures

Appendix Table A1: Robustness Check: Alternative Cohort Bands

| | (1) Years of schooling | (2) Have formal education | (3) Morbidity | (4) Disability | (5) Waged employment | (6) Self employment | (7) Days worked | (8) Hours worked |
|---|---|---|---|---|---|---|---|---|
| **Panel A. Narrower Band** | | | | | | | | |
| Cocoa Area | 1.421*** | 0.126*** | -0.009 | 0.005 | 0.024 | -0.029 | -0.075 | -0.513*** |
| | (0.215) | (0.022) | (0.022) | (0.019) | (0.021) | (0.023) | (0.097) | (0.173) |
| Target Cohort × Cocoa Area | -0.807*** | -0.055** | 0.056** | 0.047* | -0.009 | 0.008 | 0.023 | -0.035 |
| | (0.255) | (0.024) | (0.027) | (0.025) | (0.021) | (0.026) | (0.107) | (0.170) |
| Observations | 4,504 | 4,507 | 4,507 | 4,507 | 4,047 | 4,047 | 4,042 | 4,044 |
| R-squared | 0.356 | 0.333 | 0.042 | 0.102 | 0.219 | 0.106 | 0.095 | 0.130 |
| Cohort F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Region F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Outcome Mean | 4.555 | 0.498 | 0.327 | 0.214 | 0.175 | 0.730 | 5.045 | 7.136 |
| **Panel B. Wider Band** | | | | | | | | |
| Cocoa Area | 1.337*** | 0.129*** | -0.009 | -0.001 | 0.013 | -0.018 | -0.132* | -0.435*** |
| | (0.167) | (0.016) | (0.016) | (0.013) | (0.014) | (0.017) | (0.077) | (0.129) |
| Target Cohort × Cocoa Area | -1.000*** | -0.079*** | 0.050** | 0.042* | 0.007 | -0.005 | 0.077 | -0.194 |
| | (0.234) | (0.022) | (0.024) | (0.023) | (0.018) | (0.023) | (0.098) | (0.149) |
| Observations | 7,999 | 8,002 | 8,002 | 8,002 | 6,897 | 6,897 | 6,890 | 6,894 |
| R-squared | 0.375 | 0.370 | 0.041 | 0.126 | 0.220 | 0.115 | 0.095 | 0.124 |
| Cohort F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Region F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Outcome Mean | 4.973 | 0.551 | 0.313 | 0.187 | 0.165 | 0.717 | 5.008 | 7.113 |

Robustness check: Alternative Cohort Bands. The table displays the long-term effects of the cocoa swollen shoot virus disease (CSSVD) outbreak on human capital and labor market outcomes using alternative cohort bands. Cocoa area indicates respondents residing in cocoa-suitable areas. Target cohort refers to individuals born between 1935 and 1949. All models include year-of-birth and region fixed effects. Standard errors are clustered at the cohort × region level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table A2: Robustness Check: Disaggregating the Treated Cohort by Exposure Timing

| | (1) Years of schooling | (2) Have formal education | (3) Morbidity | (4) Disability | (5) Waged employment | (6) Self employment | (7) Work days | (8) Work hours |
|---|---|---|---|---|---|---|---|---|
| Cocoa Area | 1.528*** | 0.149*** | -0.008 | -0.005 | 0.017 | -0.015 | -0.121 | -0.461*** |
| | (0.187) | (0.017) | (0.017) | (0.014) | (0.015) | (0.018) | (0.079) | (0.140) |
| Only Early Childhood | -1.160*** | -0.090*** | 0.074 | 0.060 | -0.011 | 0.013 | 0.020 | 0.036 |
| | (0.373) | (0.034) | (0.045) | (0.039) | (0.027) | (0.031) | (0.195) | (0.263) |
| In-utero to Early Childhood | -1.139*** | -0.094*** | 0.051* | 0.040 | 0.006 | -0.016 | 0.089 | -0.278* |
| | (0.278) | (0.026) | (0.027) | (0.028) | (0.022) | (0.028) | (0.107) | (0.167) |
| Observations | 7,066 | 7,069 | 7,069 | 7,069 | 6,218 | 6,218 | 6,214 | 6,219 |
| R-squared | 0.363 | 0.355 | 0.039 | 0.105 | 0.221 | 0.105 | 0.094 | 0.122 |
| Cohort F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Region F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Outcome Mean | 4.986 | 0.550 | 0.309 | 0.179 | 0.170 | 0.715 | 5.024 | 7.148 |

Robustness Check: Disaggregating the Treated Cohort by Exposure Timing. The table reports the long-term effects of the cocoa swollen shoot virus disease (CSSVD) outbreak on human capital and labor market outcomes when separating exposure by timing. Cocoa area indicates respondents residing in cocoa-suitable areas. Target cohort refers to individuals born between 1935 and 1949. All models include year-of-birth and region fixed effects. Standard errors are clustered at the cohort × region level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table A3: Robustness Check: Analysis Within Cocoa-Affiliated Regions

| | (1) Years of schooling | (2) Have formal education | (3) Morbidity | (4) Disability | (5) Waged employment | (6) Self employment | (7) Days worked | (8) Hours worked |
|---|---|---|---|---|---|---|---|---|
| Cocoa Area | 1.399*** | 0.134*** | -0.012 | 0.006 | 0.018 | -0.019 | -0.106 | -0.467*** |
| | (0.197) | (0.018) | (0.017) | (0.015) | (0.016) | (0.018) | (0.081) | (0.146) |
| Target Cohort × Cocoa Area | -0.868*** | -0.058** | 0.071** | -0.005 | -0.009 | 0.006 | 0.078 | -0.122 |
| | (0.300) | (0.027) | (0.029) | (0.027) | (0.023) | (0.028) | (0.114) | (0.184) |
| Observations | 5,436 | 5,439 | 5,439 | 5,439 | 4,816 | 4,816 | 4,815 | 4,822 |
| R-squared | 0.295 | 0.258 | 0.040 | 0.105 | 0.223 | 0.125 | 0.056 | 0.119 |
| Cohort F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Region F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Outcome Mean | 5.970 | 0.656 | 0.318 | 0.206 | 0.197 | 0.727 | 4.860 | 7.131 |

Robustness Check: Analysis Within Cocoa-affiliated Regions. The table reports the long-term effects of the cocoa swollen shoot virus disease (CSSVD) outbreak on human capital and labor market outcomes within cocoa-affiliated regions. Cocoa area indicates respondents residing in cocoa-suitable areas. Target cohort refers to individuals born between 1935 and 1949. All models include year-of-birth and region fixed effects. Standard errors are clustered at the cohort × region level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table A4: Robustness Check: Falsification Check Using Post-Epidemic Birth Cohorts

| | (1) Years of schooling | (2) Have formal education | (3) Morbidity | (4) Disability | (5) Waged employment | (6) Self employment | (7) Days worked | (8) Hours worked |
|---|---|---|---|---|---|---|---|---|
| Cocoa Area | 1.229*** | 0.129*** | 0.022 | 0.016 | 0.015 | -0.014 | -0.118 | -0.470*** |
| | (0.200) | (0.018) | (0.019) | (0.016) | (0.015) | (0.018) | (0.082) | (0.137) |
| Post-epidemic × Cocoa Area | 0.005 | -0.014 | -0.052** | -0.031 | 0.008 | -0.010 | 0.053 | -0.133 |
| | (0.222) | (0.021) | (0.022) | (0.019) | (0.019) | (0.022) | (0.090) | (0.164) |
| Observations | 7,066 | 7,069 | 7,069 | 7,069 | 6,218 | 6,218 | 6,214 | 6,219 |
| R-squared | 0.361 | 0.353 | 0.039 | 0.105 | 0.221 | 0.105 | 0.094 | 0.122 |
| Cohort F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Region F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Outcome Mean | 4.986 | 0.550 | 0.309 | 0.179 | 0.170 | 0.715 | 5.024 | 7.148 |

Robustness Check: Falsification Using Post-epidemic Birth Cohorts. The table tests whether the estimated CSSVD effects persist for cohorts born after the epidemic period. Cocoa area indicates respondents residing in cocoa-suitable areas. The post-epidemic cohort includes individuals born between 1950 and 1960. All models include year-of-birth and region fixed effects. Standard errors are clustered at the cohort × region level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table A5: Migration Summary and Demographic Differences by Migration Status

**Panel A. Migration summary (full sample, $N = 7{,}070$)**

| Migration type | Count | Percent |
|---|---|---|
| Migrated across regions | 4,995 | 70.7% |
| Cocoa → Non-cocoa | 793 | 11.2% |
| Non-cocoa → Cocoa | 631 | 8.9% |

**Panel B. Migration summary (target cohorts only, $N = 1{,}906$)**

| Migration type | Count | Percent |
|---|---|---|
| Moved across cocoa/non-cocoa | 393 | 20.3% |
| Cocoa → Non-cocoa | 215 | 11.2% |
| Non-cocoa → Cocoa | 178 | 9.2% |

| | Female | Rural | Female HH head | Age of HH head | HH head | Age |
|---|---|---|---|---|---|---|
| **Panel C. Demographic differences in cocoa regions** | | | | | | |
| Cocoa → Non-cocoa | -0.154*** | -0.343*** | -0.149*** | -0.418 | 0.046 | -0.416 |
| | (0.038) | (0.037) | (0.032) | (0.519) | (0.034) | (0.302) |
| Observations | 1,202 | 1,202 | 1,202 | 1,202 | 1,202 | 1,202 |
| R-squared | 0.014 | 0.076 | 0.014 | 0.000 | 0.001 | 0.002 |
| Outcome Mean | 0.577 | 0.684 | 0.321 | 52.43 | 0.710 | 50.13 |
| **Panel D. Demographic differences in non-cocoa regions** | | | | | | |
| Non-cocoa → Cocoa | -0.044 | -0.199*** | 0.193*** | -2.305*** | 0.209*** | 0.121 |
| | (0.045) | (0.043) | (0.040) | (0.770) | (0.042) | (0.344) |
| Observations | 613 | 613 | 613 | 613 | 613 | 613 |
| R-squared | 0.002 | 0.039 | 0.049 | 0.013 | 0.036 | 0.000 |
| Outcome Mean | 0.574 | 0.729 | 0.184 | 53.51 | 0.582 | 49.99 |

Panels A and B report basic migration statistics for all respondents and for the target cohort (1935–1949). Panels C and D show demographic differences between migrants and non-migrants in cocoa and non-cocoa regions. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table A6: Robustness Checks: Accounting for Potential Bias from Migration

| | (1) Years of schooling | (2) Have formal education | (3) Morbidity | (4) Disability | (5) Waged employment | (6) Self employment | (7) Days worked | (8) Hours worked |
|---|---|---|---|---|---|---|---|---|
| Cocoa Area | 1.465*** | 0.142*** | -0.006 | -0.009 | 0.018 | -0.021 | -0.116 | -0.473*** |
| | (0.187) | (0.018) | (0.017) | (0.014) | (0.016) | (0.018) | (0.082) | (0.143) |
| Target Cohort × Cocoa Area | -1.184*** | -0.102*** | 0.047* | 0.060** | 0.003 | -0.010 | 0.077 | -0.112 |
| | (0.255) | (0.025) | (0.027) | (0.024) | (0.019) | (0.024) | (0.108) | (0.162) |
| Cocoa → Non-cocoa | -0.245 | 0.004 | 0.058*** | -0.011 | -0.047** | 0.073*** | -0.000 | -0.223 |
| | (0.233) | (0.023) | (0.020) | (0.020) | (0.022) | (0.027) | (0.088) | (0.142) |
| Non-cocoa → Cocoa | -1.044*** | -0.133*** | -0.014 | -0.006 | 0.003 | -0.071*** | 0.076 | 0.029 |
| | (0.267) | (0.025) | (0.021) | (0.016) | (0.019) | (0.024) | (0.089) | (0.174) |
| Cocoa → Non-cocoa × Target Cohort | -0.003 | -0.024 | -0.037 | 0.059* | 0.019 | -0.016 | 0.051 | 0.436** |
| | (0.442) | (0.037) | (0.039) | (0.033) | (0.031) | (0.043) | (0.165) | (0.215) |
| Non-cocoa → Cocoa × Target Cohort | 0.914** | 0.103** | 0.044 | -0.015 | 0.046 | 0.008 | 0.016 | 0.211 |
| | (0.441) | (0.044) | (0.044) | (0.038) | (0.034) | (0.042) | (0.184) | (0.292) |
| Observations | 7,066 | 7,069 | 7,069 | 7,069 | 6,218 | 6,218 | 6,214 | 6,219 |
| R-squared | 0.366 | 0.359 | 0.040 | 0.106 | 0.222 | 0.108 | 0.094 | 0.122 |
| Cohort F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Region F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Outcome Mean | 4.986 | 0.550 | 0.309 | 0.179 | 0.170 | 0.715 | 5.024 | 7.148 |

Robustness Check: Accounting for Potential Bias from Migration. The table reports the long-term effects of the cocoa swollen shoot virus disease (CSSVD) outbreak on human capital and labor market outcomes while accounting for possible bias from selective migration among the target cohorts (birth years 1936–1949). *Cocoa area* denotes areas climatically suitable for cocoa cultivation. "Cocoa → Non-cocoa" and "Non-cocoa → Cocoa" are indicators for migrants moving from cocoa to non-cocoa regions and from non-cocoa to cocoa regions, respectively. All models include cohort (year-of-birth) and region fixed effects. Controls are age of household head and indicators for female, rural residence, female-headed household, and head of household status. Standard errors are clustered at the cohort × region level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table A7: Robustness Checks: A Sample of Only Non-Movers

| | Years of schooling | Have formal education | Morbidity | Disability | Waged employment | Self employment | Days worked | Hours worked |
|---|---|---|---|---|---|---|---|---|
| Cocoa Area | 1.427*** | 0.140*** | -0.012 | 0.001 | 0.012 | -0.030 | -0.156* | -0.528*** |
| | (0.192) | (0.018) | (0.020) | (0.017) | (0.017) | (0.020) | (0.089) | (0.155) |
| Target Cohort × Cocoa Area | -1.039*** | -0.097*** | 0.059** | 0.070*** | 0.006 | -0.015 | 0.094 | -0.070 |
| | (0.274) | (0.027) | (0.027) | (0.025) | (0.020) | (0.025) | (0.112) | (0.172) |
| Observations | 5,644 | 5,645 | 5,645 | 5,645 | 4,973 | 4,973 | 4,969 | 4,976 |
| R-squared | 0.375 | 0.369 | 0.043 | 0.107 | 0.205 | 0.108 | 0.089 | 0.118 |
| Cohort F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Region F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Outcome Mean | 4.881 | 0.546 | 0.310 | 0.175 | 0.154 | 0.725 | 4.958 | 7.025 |

Robustness Check: Sample Restricted to Non-migrants. The table reports the long-term effects of the cocoa swollen shoot virus disease (CSSVD) outbreak on human capital and labor market outcomes for a sample including only non-migrant respondents. *Cocoa area* denotes areas climatically suitable for cocoa cultivation, and the target cohort refers to individuals born between 1935 and 1949. All models include cohort (year-of-birth) and region fixed effects. Controls are age of household head and indicators for female, rural residence, female-headed household, and head-of-household status. Standard errors are clustered at the cohort × region level. *** p<0.01, ** p<0.05, * p<0.1.