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

Idris Kambala [*]

DO NOT CITE

## Abstract

Beginning in the mid-1930s, cocoa farming, the primary economic activity in the forest belt of the Gold Coast (now Ghana), was severely affected by the cocoa swollen shoot virus disease (CSSVD), the most devastating virus to affect the industry. By the mid-1940s, farm output had plummeted from 30 tons to just 6 tons, with approximately 46 million trees already infected. The virus continued to spread rapidly, affecting 15 million trees annually. In addition to reduced yield, severe economic shocks, heightened stress and broader social hardships caused by the virus outbreak, the government imposed a policy of cutting down infected and nearby trees, further exacerbating the impact of the crisis. This study investigates the long-term effects of the CSSVD outbreak on the human capital and labor market outcomes of children born in cocoa-growing areas during the epidemic. Using a cohort-based analysis, I find a strong negative impact of the epidemic on human capital development but find no consistent effects on labor market outcomes. Children exposed to the epidemic in utero or early childhood experienced a three-quarter-year loss in schooling and a 6.3 percentage point (pp) decline in access to formal education. Additionally, they faced a 5 pp increase in morbidity and a 3 pp rise in early mortality in adulthood. These effects were more severe for girls, children in female-headed households, and those in urban areas. In contrast, long-term labor market outcomes, such as employment status and work periods, were largely unaffected, suggesting that individuals may have adapted over time. The results highlight how early-life negative shocks can cause lasting harm to health and education, both of which are shaped early in life, while demonstrating the potential for recovery in labor market outcomes. These findings contribute to the broader literature on early-life adversities and inform policies aimed at supporting vulnerable groups during economic crises.

**Keywords:** Gold Coast, Ghana, cocoa swollen shoot, epidemic, early-life economic shock

---

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

# I. Introduction

Since the seminal work of David J. Barker (1990), which argued that "the womb may be more important than the home," a growing body of literature has established links between early fetal development and adult-life outcomes. This idea, formalized as the Fetal Origins Hypothesis (FOH), posits that environmental conditions and exposures during the prenatal period can cause permanent changes in an individual's health, cognitive development, education, and economic well-being.[1] The premise is that fetal development is a critical window during which vital organs form, making them particularly sensitive to adverse shocks. Such shocks can produce permanent structural changes in the body, with consequences for adulthood outcomes.

The FOH has attracted significant attention in Economics, with researchers investigating how adverse early-life shocks shape adult outcomes. While the original focus was on prenatal shocks, researchers have since expanded its scope to include shocks occurring in early childhood, typically before age five. Empirical studies provide strong evidence of this hypothesis. For example, exposures to the Spanish Flu of 1918 (Almond, 2006; Lin and Liu, 2014), the Chinese Famine of 1959-61 (Almond et al., 2007; Brandt et al., 2016; Chen and Zhou, 2007), and the Dutch Winter Hunger (Ramirez and Haas, 2022) have had substantial long-term negative impacts on affected birth cohorts. An important implication of the FOH is that instead of focusing resources on treating affected adults, public policies should prioritize support for children of affected mothers early in life to improve child outcomes and enhance their future well-being as adults (Almond and Currie, 2011).

Building on this literature, this study explores a less-examined context: an agriculture-related viral disease epidemic. Unlike the well-documented effects of health-related shocks, which primarily affect human biology, less is known about how agriculture-related early-life shocks influence long-term outcomes. Agricultural epidemics could affect child development indirectly by disrupting household income, employment, food security, and parental well-being. In this study, I examine how an economic shock caused by the Cocoa Swollen Shoot Virus Disease (CSSVD) – a destructive agricultural viral disease – affected human capital formation and labor market outcomes for children born during the epidemic. By focusing on this underexplored context, this study reveals how industry-specific shocks shape child development, thereby offering new perspectives on the long-term consequences of early-life adversities.

The CSSVD outbreak was a viral disease epidemic that devastated cocoa plantations and disrupted the livelihoods of millions of cocoa farmers in the Gold Coast (modern-day Ghana). A highly destructive, mealybug-transmitted virus, the CSSVD infects cocoa trees

---

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

at every stage of development, causing yield losses of up to 70 % and killing trees within 2-3 years of infection (Ameyaw et al., 2015; Lot et al., 1991; Muller, 2008). It was first officially reported in the Gold Coast in 1936 and confirmed as a viral disease in 1939. By the time control trials began, it had already destroyed over 50 percent of the mature trees in some areas (Danquah, 2003). By the mid-1940s, the virus had spread rapidly, devastating cocoa farms and threatening the very existence of the cocoa industry.

Controlling the CSSVD proved difficult since it was a novel viral disease. After a series of field experiments, authorities determined that the only effective control method was "roguing", which involved the systematic detecting and cutting down of infected trees. Subsequently, farmers were instructed to also remove neighboring trees since adjacent asymptomatic trees were likely infected as well (Thresh and Owusu, 1986). As a result, more than 44.2 million cocoa trees were cut down from 1946 to 1955. Cocoa farmers, therefore, experienced a double economic shock: loss of yields from the virus itself and the forced removal of their cocoa trees.

This study investigates how exposure to the CSSVD outbreak affected human capital (education, health) and labor market outcomes for children born during the outbreak. The economic shock induced by the CSSVD may have affected child development through three key pathways. First, income loss from reduced cocoa production could have limited access to food and essential resources, leading to nutritional deficits that impair cognitive development, physical growth, and health (see Almond and Mazumder, 2011). Second, financial strain on households may have increased maternal stress during pregnancy, which is a risk factor for adverse birth outcomes like preterm birth and low birthweight (see Aizer et al., 2016; Case et al., 2005). Third, income constraints may have reduced parents' ability to invest in education, healthcare, and early stimulation, potentially resulting in poorer child health and delayed educational attainment (see Currie and Hyson, 1999; Royer, 2009). Therefore, I hypothesize that children exposed to the CSSVD epidemic during the early developmental window (i.e., in utero or before age five) experienced long-term deficits in education, health, and labor market participation relative to unaffected cohorts.[2]

Cocoa growing in the Gold Coast is climatically restricted to the forest belt which covers most parts of Ghana's Ashanti, Brong Ahafo, Central, Eastern, and Western regions and small portions of the Volta region.[3] Since the boundaries of cocoa-growing areas do not perfectly correspond with political administrative regions, I reconstruct a map of climatically suitable cocoa areas using historical archives. I then utilize the GPS coordinates of DHS (Demographic and Health Survey) sampling units to identify respondents residing within these areas. Using respondents' year of birth, I distinguish between treated and non-treated

---

[2] The study is not able to disentangle which of these pathways derive the main results. Hence, we estimate the "bundle effects" of the CSSVD.

[3] These regions refer to the administrative regions of Ghana as at 1990. Recent regional demarcations differ.

cohorts.

A key challenge in isolating the impact of the epidemic is identifying the treated cohorts, specifically the birth cohorts that experienced the highest exposure to the epidemic. As the viral nature of the virus outbreak was confirmed in 1939 and most farms were devastated in the 1940s, it follows that the most affected cohorts are those children born in the 1940s. A related challenge is determining the "window size" of the treated group. Using a narrow window may introduce bias in the results, while a wider window that includes "distant cohorts" could dilute the treatment effects. Furthermore, it is likely that over time, farmers would have developed new coping strategies to mitigate the impact of the economic shock. Therefore, I choose the 1940-1950 birth cohorts as my primary treated units. However, I test the robustness of my results to other treatment windows, including using a placebo-treated cohorts.

To identify the causal effect of exposure to CSSVD epidemic, I leverage variations in exposure across birth cohorts and cocoa-growing areas. The empirical strategy relies on a cohort-based Difference-in-Difference approach, which identifies the causal impact of exposure to the epidemic by interacting the indicators for treated birth cohorts and cocoa-growing areas. I include cohort fixed effects to account for common factors unique to each birth cohort. Additionally, my specification includes region fixed effects to account for any time-invariant regional factors that would potentially affect the outcome variables. When estimating the impact of the epidemic on mortality, I also control for the child's birth order.

The main findings reveal that while the CSSVD epidemic had a significant and negative impact on human capital, its effect on labor market outcomes was negligible. Specifically, children born in cocoa-growing areas during the outbreak experienced a 0.75-year loss in schooling and a 6.3 percentage point (pp) decline in formal education. Morbidity, proxied by illness or injury in the past two weeks prior to the interview, increased by about 5 pp for children affected by the epidemic. Further, early mortality, which indicates whether a child of an interviewed mother was alive or dead at the time of the interview, rose by about 3 pp for exposed children. However, being born in cocoa-producing areas during the epidemic does not have a robust effect on labor market outcomes, including wage or self-employment and work days or hours.

Heterogeneity analysis also reveals important subgroup effects. The impact of the epidemic on educational outcomes and morbidity is completely driven by girl children, suggesting that parents may have prioritized the education and well-being of boys during the crisis. Similarly, the impact of the epidemic on morbidity is driven by rural populations, and children in female-headed households experienced more severe negative effects on educational outcomes. However, the effect is not significantly different between children who grew up to become household heads and those who did not. The heterogeneity analysis also reveals no effect of the epidemic on labor market outcomes.

The null effect on labor market outcomes is puzzling, as theory suggests that educational attainment should influence labor market performance. The key implication here is that while early-life shocks have lasting effects on human capital which is usually shaped during early developmental stages, their impact on labor market outcomes appears to be negligible or short-lived. In other words, despite early disadvantages in education and health, individuals may have successfully integrated into the labor market through pathways such as job training, work experience, or new economic opportunities. These findings highlight the adaptability of individuals and labor markets, offering important insights into how early-life shocks, like pandemics and economic crises, shape long-term economic outcomes.

This study offers three main contributions. First, it is the first to investigate the long-term impacts of the CSSVD epidemic in the Gold Coast. While other cocoa diseases, such as capsids and black pods, existed, CSSVD was uniquely destructive as it threatened to exterminate the entire cocoa industry if left unchecked. On top of that, the outbreak occurred during a period when cocoa had become the Gold Coast's most valuable export, accounting for about 80 % of total exports in 1939 (Gyamera et al., 2023). Moreover, cocoa cultivation was the primary economic activity and most farmers were smallholder farmers relying on cocoa income, hence the epidemic represented a severe economic shock for farming households, with far-reaching implications for child development.

Second, to my knowledge, this is also the first study to investigate the effects of an agriculture-related virus disease that directly affected cash crop farmers. A closely related work is that of Banerjee et al. (2010) who study the long-lasting impacts of the phylloxera epidemic that destroyed 40 % of French vineyards. Their findings reveal that while the epidemic reduced adult height, it had no effect on health or life expectancy. While their study focuses on an insect pest, this study examines a viral disease, which differs in both transmission and severity. Unlike the phylloxera in France, the CSSVD not only attacked the Gold Coast's most valuable export crop, but its control also required the massive destruction of cocoa farms, further exacerbating the economic impact of the outbreak.

Third, this study contributes to the broader FOH literature on the linkages between early-childhood exposures and adult outcomes. While most of the FOH literature focuses on natural disasters, famines, and pandemics, this study explores an agriculture-related economic shock, a context that has received less attention in the literature. By focusing on the CSSVD outbreak, this study demonstrates that industry-specific shocks – particularly those related to agricultural production – can have long-term consequences for child development. This perspective broadens the FOH framework, showing that economic disruptions originating from industry-specific shocks can influence child outcomes through mechanisms beyond direct health-related shocks.

The rest of the study proceeds as follows. In the next section I provide an account of cocoa production and the onset of the CSSVD in the Gold Coast. In section III. I discuss the empirical strategy and the data, and in section IV. I report the main empirical estimates plus the heterogeneous effects and robust checks. Section V. concludes the study.

## 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 he had acquired during his travels abroad. Initially, cocoa farming remained an unpopular agricultural venture until about the mid-to-late 1880s when it started to gain wider attention. The first recorded export of cocoa of about 121 lbs occurred in 1885, and it was not until six years later in 1891 that another, smaller shipment of 80 lbs was made. The industry's spectacular growth took place 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 firmly established itself as the world's leading cocoa producer.

Cocoa farming was a profitable venture and attracted the attention of natives who hitherto focused on other crops such as coffee and oil palm. All that was needed to start a cocoa farm was a fertile land with some intense labor work at the early stages. It requires comparatively less maintenance, and once it attains the fruit-bearing age, it can produce recurring yields for more than four decades (Gyamera et al., 2023). Hence, cocoa farming became the primary economic activity in the climatically suited cocoa areas of the country.

Unfortunately, the greatest threat to the cocoa boom was already on the horizon. Beginning in the 1930s, cocoa farmers started observing escalating dieback on their cocoa trees. No one knew what was causing it until a farmer, in 1936, brought a sample of infected cocoa tree to the department of agriculture for further investigation. The farmer showed to the pathologist at the department of agriculture the swollen branches of the cocoa tree. The pathologist, W.F. Steven, observed that the distended branches could be a result of an infectious disease, an observation later confirmed by field studies. 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 of its leaves, loses the shape of its fruit-bearing spherical pod and gradually dies. Because the most characteristic symptom of the disease was swollen stem and shoot, it was given the name the "swollen shoot" virus disease.

The CSSVD was an entirely new disease, presenting a challenging puzzle for effective control. In the early days, Dade, a mycologist, thought that the disease was caused by the lack of enough shading for cocoa trees. Hence, the government instructed cocoa farmers to plant cocoa under a canopy of trees that provided dense shading. Farmers who violated this practice were to be punished. This was the practice until in 1939 when it was finally confirmed that the disease was caused by a virus.

To control the spread of the virus, a new control method called roguing, which involved the careful detection and removal of infected cocoa trees from the farms, was introduced. While this was a better method, it was not entirely effective because seemingly healthy trees bordering diseased trees were likely also infected and would show disease symptoms only in a matter of time. Accordingly, it was later recommended that all symptomatic and adjacent asymptomatic trees be cut out: all trees within 5, 10 and 15 meters of outbreaks 1-10, 11-100 and 100+ of infected trees were to be removed, respectively (Thresh et al., 1988).

In the early years of the operation, no records were kept on the number of trees cut. However, data from the mid-1940s provides insights into the scale of diseased tree removal. Figure 1 illustrates the number of diseased trees cut annually from 1945/6 to 1955/6. Between 1945 and 1950, the total number of trees cut ranged from 52,000 to 1 million. But after 1951, this figure surged dramatically, reaching a peak of 5.5 million in 1954/5. A more nuanced observation emerges from examining the total number of trees cut per acre. 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, a substantially higher number of trees/acre must have been cut during the early years of the operation.

The cutting-out method stirred strong public discontent because farmers would prefer reaping smaller harvest from infected farms than face a complete loss of income that would result from destroying the entire farm. The income loss was even worse for settler farmers since their landlords could take back the lands at the end of the growing season (Dzahini-Obiatey et al., 2010). The cutting out method was also unpopular because cocoa farming is a long-term investment. It demands considerable investments in land, labor and credit. Moreover, seedlings need about seven uninterrupted years to start bearing fruits. The public discontent and non-cooperation stymied government's control efforts until 1946 when a legislative instrument was enacted to make cutting-out method mandatory. Farmers would be arrested and prosecuted for non-cooperation.

By the mid 1940s the virus had spread wild and caused an alarming devastation to large expanses of cocoa farms. Between 1939 and 1944, about two-thirds of the trees planted over the period 1904-1914 were devastated by the virus. Output per farm dropped 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: it proceeded at 15 million trees per year, and it was therefore estimated that the cocoa industry would disappear in the next 20 years if the virus were left unchecked (Danquah, 2003). To placate public discontent, in 1948 the colonial government sought independent expert opinion from the Food and Agriculture Organization (FAO) of the United Nations who sent three plant pathologists to study the disease. The experts' recommendation confirmed the earlier proposition of eradication of infected and nearby trees – and unless such was done the complete extinction of the industry was only a matter of time. Thus, the compulsory cutting-out method continued.

In this study, I investigate the effects of the CSSVD epidemic, focusing on its long-term effects on children conceived and born during the outbreak. The epidemic led to substantial economic losses for cocoa farmers due to the destruction of cocoa trees and loss of farm income.[4] Additionally, reduced family resources may have limited parental investments in children's education, healthcare, and early stimulation, exacerbating developmental impacts. Beyond the economic loss, the epidemic may have triggered maternal stress in pregnant women, which has been linked to long-term mental health disorders in children, including anxiety, depression, and ADHD (see Persson and Rossin-Slater, 2018). Taken together, these pathways suggest that children born in cocoa-growing areas during the peak of the epidemic experienced a combination of nutritional deprivation, maternal stress, and reduced parental investments – all of which have been shown to produce lasting developmental impacts. Accordingly, my main hypothesis is that the CSSVD epidemic had persistent adverse effects on the education, health, and labor market outcomes of children born in cocoa-growing areas during the outbreak.

While the viral infection and the eventual cutting out of trees both represent significant economic losses, I am not able to disentangle their individual effects. My empirical strategy is primarily structured to pick up the effects of the viral infection, which imposed uncompensated economic losses and stressors on cocoa-farming households. As explained below, when the cutting out began, monetary rewards were given to farmers whose trees were cut. While such compensation may have mitigated the developmental impacts of the epidemic, my analysis does not identify the extent to which these payments offset the overall impact. To address these countervailing factors, my treated window focuses on the period corresponding to the height of the virus's spread, as well as the period when compensatory payments were least likely to affect child development. Lastly, although the epidemic may have affected developmental and adult outcomes through multiple pathways, my analysis does not identify the relative contributions of these pathways to the observed effects. Hence, I estimate the total (or 'bundle') effect of exposure to the CSSVD epidemic, capturing the combined impact of multiple, overlapping mechanisms.

---

[4] A large literature also shows that parental income and family socioeconomic status are key determinants of children's developmental outcomes (see Akee et al., 2010; Berger et al., 2009; Blau, 1999; Currie, 2009; Dahl and Lochner, 2012; Duncan et al., 2011).

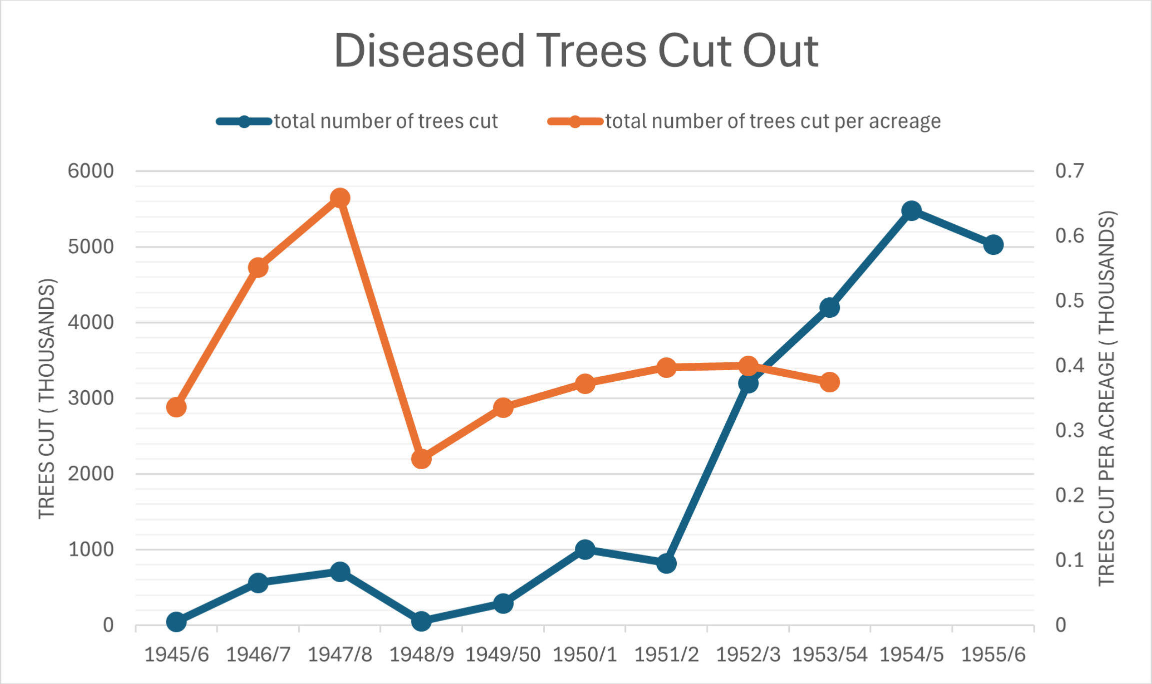

Figure 1: The figure shows the number of diseased trees cut annually from 1945/6 to 1955/6. Data from (Dickson, 1960)

## III. Data and Methodology

In this section I discuss the treatment window, the data, and the empirical framework.

### III.I Identifying the Treated Units

The key challenge in isolating the effects of the epidemic is identifying the treated cohorts. Although the viral nature of the disease was confirmed in 1939, the economic consequences were unlikely immediate. Indeed, the virus spread rapidly in the ensuing decade, inflicting its most severe impacts during the 1940s. In addition, wartime neglect further exacerbated the situation, as the government was slow to implement control measures during World War II. As a result, cocoa yields declined significantly, exposing cocoa-farming households to severe economic hardship. Given this context, I identify children born in the 1940s as the most affected cohorts.

A related challenge is the size of the "treatment window." On one hand, a narrow window may introduce bias if it excludes children who were plausibly exposed to the epidemic. On the other hand, a wider window that includes "distant cohorts" could dilute the effects I aim to isolate. To address this tradeoff, I define the 1940–1950 period as the primary exposure window. This period captures the years when the disease had its most devastating impact on cocoa farms – a time when government compensation schemes were either minimal or non-existent.

The rationale for selecting this window requires further institutional elaboration. Early in the epidemic, compensatory payments to cocoa farmers were minimal and unlikely to offset the severe economic shock. Initially, farmers were only compensated for neighboring healthy trees that were cut as part of control measures to prevent the virus's spread. It was not until 1947 that a legislative instrument was introduced, mandating the compulsory removal of infected trees while providing more generous compensation for farmers. However, it was in June 1951 that the government announced the "New Deal for Cocoa", which offered the most substantial and generous compensatory payments for farmers. This timeline is critical because it marks a shift in the economic conditions of cocoa-farming households. Before 1951, farmers bore the full economic brunt of the epidemic, but after 1951, government support through cash transfers could have mitigated the impact on children's developmental outcomes.

By focusing on the 1940–1950 window, I avoid contamination from these large compensatory payments. This ensures that observed effects are driven by the CSSVD epidemic rather than later government interventions. Additionally, this window reduces the likelihood that observed effects are biased by household coping strategies, such as farm diversification or changes in labor supply. In the early years of the epidemic, farmers were likely caught off guard by the sudden loss of income, whereas later cohorts may have had time to adjust to the economic shock. Thus, the 1940–1950 window provides a clean identification strategy that captures the epidemic's "bundle effect," reflecting the combined impact of economic loss, reduced household resources, and maternal stress.

### III.II Data

The main data used to identify the effect of the epidemic is the standard DHS (Demographic Health Survey) conducted in Ghana in 1993.[5] By 1993, the treated cohort (children born between 1940 and 1950) would have been between 43–53 years old. I restrict the sample to respondents aged 25–73, which includes individuals born between 1920 and 1968. I also exclude foreign-born respondents.

DHS is a nationally representative survey conducted primarily in low- and middle-income countries. With support from the USAID, the DHS 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/AID. The DHS also includes information on other socioeconomic indicators such as domestic violence, education, and labor market variables. The survey is conducted at the household level and primarily targets women of reproductive age (15-49). Although men usually make up a smaller portion of the sample, the recode I used has no disproportionate gender representation.

To streamline analyses, the survey data is organized into several recodes. For this study, I use the Household Member Recode (HHMR) and the Birth Recode (BR) for different analyses. The HHMR provides information on household members' educational attainment, morbidity, work hours and days, and waged or self employment. For educational attainment I look at years of schooling (in single years) and access to formal education (respondent has formal education or not). I proxy morbidity by whether the respondent reported being ill or injured in the previous two weeks. Lastly, waged employment refers to whether respondent works for someone else for wages or salaries, while self employment indicates that respondents work for themselves.

From the BR, I examine whether children born during the epidemic in cocoa-growing areas are more likely to experience early mortality. Interviewers collect information on each birth reported by a mother, indicating whether a child was alive or deceased at the time of the interview. I create a binary variable equals one if a child was deceased (at the time of the interview) and zero if alive. This allows me to assess the relationship between exposure to the epidemic and child mortality.

---

[5] The first standard DHS in Ghana was conducted in 1988. However, this data lacks GPS and locality information for mapping.

Since cocoa and non-cocoa areas are not administratively defined, I map respondents on the map of the Gold Coast (Ghana) using the GPS coordinates of their respective primary sampling units. By projecting their GPS coordinates onto the map shown in figure 2 I can effectively determine whether a respondent resides in a cocoa-growing area or not. I later assess the sensitivity of my estimates by considering the possibility of migration since it is likely that children born in cocoa areas may have relocated to non-cocoa areas later in life; or conversely, individuals from non-cocoa areas might have moved into cocoa areas. Both scenarios would bias the estimates.

The summary statistics are presented in table 1. The statistics for individual household members (male and female from the HHMR) are displayed in the top panel, and those for children (BR) are shown in the bottom panel. The total number of observations from the HHMR sample ranges from 6,479 to 7,427. The missing values come from the labor market outcomes which is likely due to recall problems or the exclusion of people not in the universe. In contrast, the BR has a larger sample size (12,193) since the unit of analysis is a child born to an interviewed woman.

Household members in the HHMR dataset have an average age of 41.5 years and approximately five years of schooling on average. The majority of respondents, 53 percent, have had some form of formal education. In terms of morbidity, more than three out of ten people reported being ill (or injured) in the past two weeks. Most of respondents (62 percent) are household heads, and the average age of these householders is 45 years. Although more than half of the sample is female, only 27 percent of householders are female. The data also shows that a substantial proportion (46 percent) of respondents reside in cocoa-growing areas. Rural respondents make up a disproportionately higher share (68 percent) of the sample. Lastly, exactly one-fifth of the sample consists of the treated cohort, a significant portion.

Turning to the BR in the bottom panel, the average age of a child in the sample is about 9 years while the minimum and maximum ages are zero and 35.[6] As at the time of the survey about 15 percent of children born to surveyed woman had passed away. Nearly half of the children in the sample also live in cocoa areas and about one-fourth of them belong in the treated group.

In both panels, some noticeable disparities emerge between the treated and untreated cohorts with respect to some related outcomes. The treated cohorts have lower years of schooling, less likely to get a formal education and show higher morbidity and mortality rates. However, the treated cohorts tend to work longer hours and are slightly more likely to be waged employed. Conversely, the untreated cohorts tend to be more self employed. My empirical strategy interacts indicators of the treated cohorts and cocoa areas to determine

---

[6] For deceased children I input their age at death.

the additional impact of the epidemic on these disparities.

### III.III Empirical Framework

My primary objective of is to estimate the causal impact of exposure to the CSSVD on long-term human capital and labor market outcomes. Specifically, I examine whether children born in cocoa-growing areas during the outbreak exhibit poorer adult outcomes relative to their counterparts born in unaffected areas. The key outcomes I analyze include educational attainment (years of schooling and access to formal education), health outcomes (early mortality and morbidity), and labor market outcomes (waged or self-employment and days or hours worked).

To identify the long-term effects of CSSVD exposure, I employ a cohort-based Difference-in-Differences (DID) approach, estimating the following specification:

$$y_{icr} = \sum_{c=1920}^{1968} \gamma_c \, cohort_c + \alpha \, Cocoa\_Area_i + \mathbf{FE}_r + \beta \, TreatedCohort_c \cdot Cocoa\_Area_i + \varepsilon_{icr} \quad (1)$$

Where $y_{icr}$ is the outcome of individual $i$, born in cohort $c$ and resides in region $r$. $cohort_c$ represents a full set of cohort dummies for birth cohorts ranging from 1920 to 1968. The cohort fixed effects ensure that comparisons are made within birth cohorts, not across them. For example, educational attainment may have improved over time due to nationwide investments in education, and these trends are captured by cohort fixed effects. By including these cohort effects, I isolate the specific impact of being born during the CSSVD epidemic from broader birth cohort trends.

The term $Cocoa\_Area_i$ is a binary variable equals to 1 if individual $i$ resides in an area that is climatically suitable for cocoa production, as shown in Figure 2, and 0 otherwise. The climatically suitable areas include most of Ghana's Ashanti, Central, Eastern and Western regions, as well as parts of the Volta region. As the figure shows, the boundaries of the climatically suited areas do not perfectly match with those of the administrative regions. Focusing on the climatically suited areas allows me to closely isolate the effect of the epidemic. The coefficient $\alpha$ on $Cocoa_Area_i$ captures the average difference in outcomes between individuals born in cocoa-growing and non-cocoa-growing areas, regardless of birth cohort.

$\mathbf{FE}_r$ represents region fixed effects, capturing any time-invariant region-specific factors that could affect child development. For example, some regions might have better access to education or health services, independent of the CSSVD epidemic. In addition, note that most of the administratively defined cocoa-producing regions include both cocoa-growing and non-cocoa-growing areas. Therefore, by including $Cocoa\_Area_i$ and $\mathbf{FE}_r$ simultaneously, I ensure that I am comparing children born at the same time in the same administrative region but with different exposure to the epidemic, thereby controlling for confounding factors at the regional level.

The key variable of interest is the interaction term $TreatedCohort_c \cdot Cocoa_Area_i$, which identifies children born in cocoa-growing areas during the epidemic years. The coefficient $\beta$ captures the causal impact of being born in a cocoa-growing area during the epidemic, relative to being born outside a cocoa-growing area or outside the treated cohort window. Lastly, to account for potential within-group correlation, I cluster standard errors at the cohort × region level, allowing for arbitrary correlation of errors within both cohorts and regions.

Table 1: Summary Statistics

| | (1) N | (2) min | (3) max | (4) mean | (5) sd | (6) N | (7) mean | (8) N | (9) mean |
|---|---|---|---|---|---|---|---|---|---|
| | Household Member Recode | | | | | | | | |
| | Whole Sample | | | | | Treated | | Untreated | |
| Years of schooling | 7,424 | 0 | 20 | 4.805 | 5.143 | 1,487 | 4.111 | 5,937 | 4.978 |
| Formal education | 7,427 | 0 | 1 | 0.531 | 0.499 | 1,488 | 0.444 | 5,939 | 0.553 |
| Morbidity | 7,427 | 0 | 1 | 0.315 | 0.465 | 1,488 | 0.336 | 5,939 | 0.310 |
| Days worked | 6,472 | 0 | 7 | 5.000 | 1.792 | 1,363 | 5.099 | 5,109 | 4.974 |
| Hours worked | 6,479 | 0 | 24 | 7.103 | 2.902 | 1,363 | 7.261 | 5,116 | 7.061 |
| Waged employment | 6,479 | 0 | 1 | 0.165 | 0.371 | 1,363 | 0.186 | 5,116 | 0.159 |
| Self employment | 6,479 | 0 | 1 | 0.723 | 0.447 | 1,363 | 0.712 | 5,116 | 0.726 |
| Age | 7,427 | 25 | 73 | 41.51 | 12.80 | 1,488 | 47.99 | 5,939 | 39.88 |
| Cocoa area | 7,427 | 0 | 1 | 0.464 | 0.499 | 1,488 | 0.451 | 5,939 | 0.468 |
| Treated/Untreated cohort | 7,427 | 0 | 1 | 0.200 | 0.400 | - | - | - | - |
| Female | 7,427 | 0 | 1 | 0.561 | 0.496 | 1,488 | 0.564 | 5,939 | 0.560 |
| Head of household | 7,427 | 0 | 1 | 0.622 | 0.485 | 1,488 | 0.651 | 5,939 | 0.615 |
| Age of household head | 7,427 | 17 | 98 | 45.32 | 13.68 | 1,488 | 50.88 | 5,939 | 43.93 |
| Female head of household | 7,427 | 0 | 1 | 0.268 | 0.443 | 1,488 | 0.247 | 5,939 | 0.274 |
| Rural | 7,427 | 0 | 1 | 0.675 | 0.468 | 1,488 | 0.686 | 5,939 | 0.672 |
| | Birth Recode | | | | | | | | |
| Mortality | 12,193 | 0 | 1 | 0.150 | 0.357 | 3,012 | 0.186 | 9,181 | 0.139 |
| Cocoa area | 12,193 | 0 | 1 | 0.472 | 0.499 | 3,012 | 0.445 | 9,181 | 0.481 |
| Treated/Untreated cohort | 12,193 | 0 | 1 | 0.247 | 0.431 | - | - | - | - |
| Female | 12,193 | 0 | 1 | 0.485 | 0.500 | 3,012 | 0.482 | 9,181 | 0.486 |
| Age | 12,193 | 0 | 35 | 8.956 | 7.325 | 3,012 | 13.69 | 9,181 | 7.401 |
| Head of household | 12,193 | 0 | 1 | 0.320 | 0.467 | 3,012 | 0.323 | 9,181 | 0.319 |
| Age of head of household | 12,193 | 18 | 95 | 43.56 | 10.96 | 3,012 | 51.59 | 9,181 | 40.93 |
| Female head of household | 12,193 | 0 | 1 | 0.349 | 0.477 | 3,012 | 0.337 | 9,181 | 0.353 |
| Rural | 12,193 | 0 | 1 | 0.716 | 0.451 | 3,012 | 0.715 | 9,181 | 0.716 |

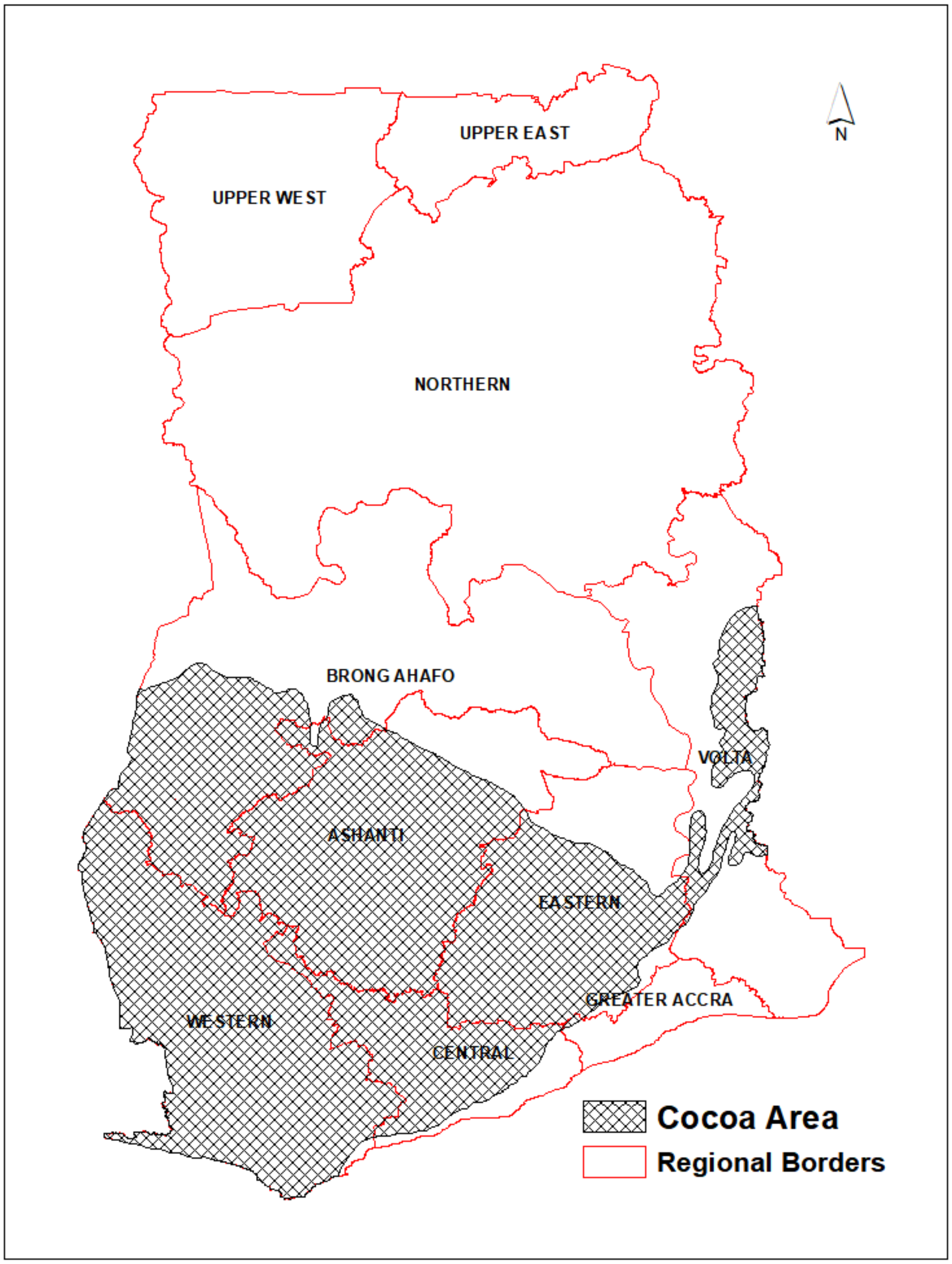

Figure 2: The figure shows the climatically suited cocoa-growing areas and the ten regions of Ghana

## IV. Results

In this section I report and discuss the empirical estimates of the CSSVD exposure. The main results are report in tables 2 and 3 for human capital and labor market outcomes respectively. The odd-numbered columns contain no controls aside from cohort (year of birth) fixed effects and region fixed effects. The common controls in the even-numbered columns are age of household head, and indicators for female, household head, female head, and rural residents. For the mortality variable (columns 7 & 8 of table 2), I also add indicators for the child's year of birth. I focus my analysis on the conditional estimates in the even-numbered columns.

The results in table 2 show that the swollen shoot disease had a significant negative effect on human capital. Although cocoa-growing areas had higher human capital in the form of greater years of schooling and access to formal education, being born in these areas during the epidemic period is associated with lower human capital outcomes. The estimates indicate that the epidemic reduced years of schooling by three-quarter of a year, corresponding to a 15.6 percent drop in average years of schooling. The epidemic also decreased the probability of having formal education by 6.3 percentage points (pp), reducing the average by about 12 percent.

For morbidity (column 6), the estimates show no differences between cocoa and non-cocoa areas. However, the epidemic increased morbidity rate in cocoa areas. Specifically, being born in cocoa areas during the epidemic years increased one's morbidity rate by 5 pp. The conditional estimates in column (8) also indicate that while mortality was significantly lower in cocoa-growing areas, the epidemic increased the mortality rate by 3 pp for exposed children.

The above results can further be discussed in terms of the *combined* effects. The coefficient of "cocoa area" in column 2 indicates that individuals in cocoa-growing areas had 1.30 additional years of schooling compared to those outside these areas. However, for children born during the epidemic years in cocoa areas, this gap reduced to a little over one-half additional years of schooling (0.55 = 1.30 - 0.81), representing a 58 percent reduction in the gap between cocoa and non-cocoa areas. Similarly, the epidemic reduced the gap in access to formal education between cocoa and non-cocoa areas from 13 pp to 6.7 pp. For morbidity, which already shows no significant difference between cocoa and non-cocoa areas in the absence of the outbreak (columns 5 and 6), the epidemic had a much larger effect. Specifically, starting from a base of zero, the epidemic increased morbidity rate by 0.05 points for children born during the epidemic in cocoa areas. Lastly, mortality shows an even larger shift. Cocoa areas had a 2.3 pp lower mortality rate compared to non-cocoa areas. However, the epidemic flipped this gap such that, for treated cohorts, cocoa areas now had 0.7 pp (-0.023 + 0.030) higher mortality rate relative to non-cocoa areas.

Turning to table 3, the results suggest that the epidemic did not have a significant impact on labor market outcomes. None of the key coefficients is statistically significant at conventional levels. Additionally, there are no significant differences between cocoa and non-cocoa areas in terms of waged or self employment. However, one notable finding is that individuals in cocoa areas worked fewer days and hours compared to those in non-cocoa areas.

This finding highlights one important implication about the long-term impacts of early life shocks. It suggests that while early-life shocks undoubtedly leave lasting effects on outcomes like educational attainment, mortality and health status that are shaped during critical developmental stages, their effects on labor market outcomes may be minimal or fade over time since the labor market offers flexibility for adaptability through mechanisms like job training, skill acquisition and work experience. In other words, although education is a strong determinant of labor market success, the flexibility in the labor market allows individuals to mitigate or overcome the effects of early-life disadvantages.

## IV.I Heterogeneous Effects

I examine whether the long-term effects of the epidemic differ by demographic subgroups including gender, type of place of residence, gender of household head, and householder status. To do this, I add a new interaction term which represents the multiplicative term of each demography variable and the original $cocoa\_area \times treatedcohort$ interaction term. The coefficient of this new interaction term indicates whether the effect of the epidemic varies by a given subgroup. The results are reported in tables 4–7.

### IV.I.I Gender-Specific Effects

Firs, I look at whether the results vary by gender. Accounting for gender-specific effects (table 4), the effect of the epidemic is minimal and not statistically significant for years of schooling, access to formal education, and morbidity. Only the coefficient of mortality remains unchanged and still statistically significant. The new interaction term produces significant results for education levels, access to education, and morbidity, suggesting that the negative impact of the epidemic on human capital was entirely driven by female children. Specifically, female children born in cocoa areas between 1940 and 1950 experienced an additional 1.7 years loss of schooling, a 16.3 pp decline in access to formal education, and 9.5 pp increase in morbidity compare to male children in similar conditions. A plausible explanation would be that parents started to prioritize the education and wellbeing of boys (to the detriment of girls) during the epidemic years.

The labor market outcomes also show an interesting set of results. Accounting for gender-specific effects, the epidemic had a positive effect on waged employment and, conversely, a negative impact on self employment. However, for female children, the positive impact of the epidemic on waged employment is 12 pp less intense whereas its negative impact on self employment is less severe by 14 pp. As shown above, a likely explanation for these findings is that the epidemic disproportionately impacted female educational outcomes, hindering their entry into the labor market and forcing them into self employment.

### IV.I.II Type of Place of Residence

In table 5, I examine whether the results are driven by rural and urban dynamics. The adverse impact of the epidemic on educational outcomes is still significant regardless of type of place of residence. However, residing in a rural area significantly reduced the negative impact of the epidemic on access to formal education. The impact of the epidemic on morbidity was, however, completely driven by rural residents. The labor market outcomes produce no statistically significant results for the key variables.

### IV.I.III Household Head

I also look at whether the results are driven by householder (the household head) and non-householder differences. The results, reported in table 6, indicate that except for years of schooling which shows that being a household head lessened the impact of the epidemic, the effect of the epidemic was not largely driven by householder and non-householder differences. The labor market variables do not also produce consistent estimates.

### IV.I.IV Gender of Household Head

Lastly, I investigate whether the results would differ by gender of a household head. The results, displayed in table 7, suggest that the negative effect of the epidemic on educational outcomes was driven entirely by members in female-headed households. For treated individuals in female-headed households, the reduction in years of schooling is greater by 1.3 years compared to those in male-headed households in cocoa areas. This finding highlights the heightened vulnerability of female-headed households during crisis. However, the effect of the epidemic on morbidity and mortality is not driven by the gender of household heads.

## IV.II Robustness Checks

In this section, I run some robustness checks to examine the sensitivity of my results.

### IV.II.I Placebo Treated Cohorts

I have argued previously that since the peak of the epidemic occurred between 1940 and 1950, only children born during this period would be distinctly affected. In this robustness exercise, I attempt to falsify this hypothesis by examining different cohorts who had minimal exposures to the epidemic. Specifically, I focus on cohorts born in the decade immediately following the height of the epidemic, which consist of children born between 1951 and 1960.

Although the CSSVD remained prevalent even after 1950, there are two primary reasons why its impact would have diminished by that time. First, it is reasonable to assume that, by this time, cocoa farmers had adopted various coping strategies to mitigate the impact of the epidemic. Second, as stated before, starting from the 1950s the government began implementing the most generous compensatory payments to cocoa farmers to alleviate the impact of the epidemic. For example, the newly introduced "New Deal for Cocoa," in June 1951 provided the most generous compensation packages to affected farmers to date. Therefore, I anticipate that individuals born between 1951 and 1960 were largely unaffected by the epidemic.

The results of this falsification exercise are displayed in table A1. The results confirm that the birth cohorts of 1951–1960 were not systematically affected by the epidemic. Almost all the key point estimates are not statistically different from zero. Only the morbidity variable is statistically significant, but with a counterintuitive sign suggesting that the epidemic reduced morbidity. In sum, the findings suggest that only the cohorts born during the height of the epidemic were adversely affected in adulthood.

### IV.II.II Placebo Cocoa Regions

I conduct a similar exercise where I replace the "cocoa areas" indicator with a dummy representing regions that are climatically unsuited for cocoa cultivation. This sensitivity analysis allows me to verify that only children born in cocoa areas were adversely affected by the epidemic.

Administratively, Ghana had ten regions as of 1993, with six of them – Ashanti, Brong Ahafo, Central, Eastern, Western, and Volta regions – being suitable for cocoa production. The remaining four regions – Greater Accra, Northern, Upper East, and Upper West – are not climatically suited for cocoa cultivation. I assign these four regions as the "cocoa regions" and rerun the regression.[7] In this specification, I interact the placebo cocoa regions with the treated cohort indicator.[8] I expect the point estimates to be (close to zero and) statistically insignificant which would suggest that the impact of the epidemic was not driven by broader trends that also affected non-cocoa areas during the epidemic years.

The estimates are reported in appendix table A2. As previously uncovered, non-cocoa regions generally perform worse than cocoa regions. However, the results do not provide

---

[7] It is important to note that while these four non-cocoa regions are administratively defined, they are different from areas that are not locally climatically suited for cocoa cultivation. The reason is that most of the cocoa-producing regions also have non-cocoa areas.

[8] Note that the inclusion of the non-cocoa regions indicator prevents me from adding region fixed effects in the specification.

robust evidence that the epidemic differentially affected children born in non-cocoa areas. Only the coefficient estimate of morbidity is negative and marginally significant. This is not surprising since the results already indicate that morbidity is lower in non-cocoa areas. Overall, the results suggest that no broader or other factors differentially (dis)favored non-cocoa areas in terms of human capital and labor market outcomes during the height of the epidemic.

### IV.II.III A Discussion on Migration

This section discusses whether selective migration may have biased my results. Selective migration could bias my estimates upward or downward depending on the general defining characteristics of immigrants and out-migrants. If, for example, out-migrants (from cocoa areas) were the most adversely affected families during the epidemic, the estimates would be downward bias. I discuss the potential bias in my results that may arise from such possible scenarios.

The HHMR from the DHS includes a question asking respondents to provide information about the region they originally moved from if they were not "currently" residing in their region of birth.[9] Using this question, I examine the demographics of respondents moving in and out of cocoa-producing regions. I also test if my estimates are affected by these migration patterns. This question also allows me to examine the ultimate educational and labor market outcomes (in destination places) of exposed individuals that moved outside their birth regions.

The HHMR depicts a considerably high migration rate in my sample (those born between 1920–1968). About 70.8 percent of the respondents (5,256 out 7427) reported moving from a different region to their current place of residence. However, for this to significantly bias my results, it must be the case that people moved en-masse between cocoa and non-cocoa regions since the exposure to the treatment is concentrated in the former. This type of movement in my sample is extremely low, however. Only 11 percent of respondents moved from cocoa to non-cocoa regions, and an ever lower percent (9 percent) relocated from non-cocoa to cocoa regions. The higher out-migration rate from cocoa regions suggests that the virus outbreak in these areas acted as a push factor.

Examining some demographic variables also reveals some distinct differences between these two sets of migrants. While both groups have similar average age of 41 years, those moving from non-cocoa to cocoa regions are more likely to be female, household heads, rural residents (destination place), and female household heads than those moving from cocoa to non-cocoa regions. These differences are likely to bias my results, although the direction of bias may be indeterminate. However, given that a larger proportion of the non-cocoa-to-

---

[9] The BR has no such question so I am not able to carry out such exercise for mortality.

cocoa movers are female and reside in rural areas, both of which historically faced greater barriers to education and formal employment, my main results are likely to be downward biased.

In what follows, I run a regression that accounts for the possible effect of selective migration. In this model, I add indicators representing these two groups of migrants and interact them with the indicator for treated birth cohorts. The coefficients of the interactions will indicate how human capital and labor market outcomes responded for *treated individuals* that moved between cocoa and non-cocoa regions relative to those who did not move. The results are displayed in table A3.

As the results show, accounting for the potential effects of migration does not materially change the main results. The estimated coefficients are not very different from those of the main results in tables 2 and 3. The epidemic continues to have a statistically significant negative effect on educational outcomes, reducing years of schooling by .85 years and access to formal education by 8 pp. However, the coefficient for morbidity is no longer statistically significant.

Analyzing the indicators for migrant types also reveals some interesting patterns. Compared to non-migrants, those that moved from cocoa to non-cocoa areas had higher morbidity rates, less likely to be waged employed, and worked fewer hours. In other words, out-migrants from cocoa regions tended to be among the worse-off. On the other hand, when compared to stayers, migrants from non-cocoa regions that settled in cocoa regions exhibited comparatively worse educational outcomes and were less likely to be self employed. Together, these patterns suggest selective migration.

Looking at the interactive terms, only the coefficients for access to formal education and work hours are statistically significant for migrants moving from cocoa to non-cocoa regions. These interesting results deserve further discussion. The negative and significant coefficient for access to education implies that migrating out of the cocoa regions did not shield individuals from the negative effects of the epidemic on educational outcomes: the educational disadvantages faced by the treated cohort in cocoa regions persisted even after relocating. The lack of significance for years of schooling might suggest that, although access to education itself was constrained for this group, those who did gain access (even after migrating) pursued similar levels of schooling as their non-migrant cohorts in the cocoa regions.

The fact that the coefficient of work hours is positive and statistically significant for the treated cohorts that moved from cocoa to non-cocoa regions suggests that relocating outside of cocoa regions was associated with better labor market outcomes in terms of working hours. One implication of this finding is that treated individuals that moved from cocoa to non-cocoa regions did so for better labor market opportunities, possibly in jobs and sectors in non-cocoa regions where either labor demand was higher or that required longer working hours. Another implication is that, after migrating, working for long hours was a coping mechanism to recover from the economic shocks they or their families experienced in the cocoa regions during the epidemic years.

In sum, while selective migration is evident, migration itself does not systematically bias my main results. Migration outside cocoa regions allowed the affected individuals to increase their labor market participation chances in non-cocoa regions. It is notable, however, that while migration out of cocoa regions for the treated cohort is associated with more hours worked, it is also linked with lower access to formal education. This suggests that, for the treated cohorts that moved outside of cocoa regions, education and labor market decision were driven by different factors post migration. Individuals may have prioritized labor market participation over further education after moving to non-cocoa regions, or they may have faced trade-offs between schooling and working hours.

Table 2: The Long-term Effects of Cocoa Swollen Shoot Disease on Human Capital

| | (1) years of schooling | (2) years of schooling | (3) formal education | (4) formal education | (5) morbidity | (6) morbidity | (7) mortality | (8) mortality |
|---|---|---|---|---|---|---|---|---|
| cocoa area | 0.935*** | 1.297*** | 0.104*** | 0.129*** | 0.006 | 0.000 | -0.017 | -0.023** |
| | (0.213) | (0.176) | (0.018) | (0.016) | (0.018) | (0.018) | (0.011) | (0.011) |
| treated cohort × cocoa area | -0.763*** | -0.750*** | -0.064** | -0.063*** | 0.048* | 0.050** | 0.031** | 0.030** |
| | (0.283) | (0.257) | (0.025) | (0.024) | (0.026) | (0.025) | (0.015) | (0.015) |
| Constant | 4.439*** | 8.013*** | 0.489*** | 0.739*** | 0.308*** | 0.148*** | 0.155*** | 0.130*** |
| | (0.105) | (0.367) | (0.009) | (0.037) | (0.009) | (0.039) | (0.006) | (0.021) |
| | | | | | | | | |
| Observations | 7,424 | 7,424 | 7,427 | 7,427 | 7,427 | 7,427 | 12,191 | 12,191 |
| R-squared | 0.261 | 0.374 | 0.297 | 0.365 | 0.033 | 0.043 | 0.039 | 0.042 |
| 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 | No | Yes | No | Yes |
| Outcome Mean | 4.805 | 4.805 | 0.531 | 0.531 | 0.315 | 0.315 | 0.150 | 0.150 |

The table displays the long-term effects of the swollen shoot cocoa disease on human capital. The cocoa area refers to places climatically suited for cocoa cultivation and the treated cohort captures all births between 1940–1950. All models include cohort (year of birth) fixed effects and region fixed effects. The controls are age of household head and indicator variables for female, rural, female head of household, and head of households. Models (7) and (8) also include child's year of birth fixed effects. 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 Cocoa Swollen Shoot Disease on Labor Market Outcomes

| | (1) waged employment | (2) waged employment | (3) self employment | (4) self employment | (5) work days | (6) work days | (7) work hours | (8) work hours |
|---|---|---|---|---|---|---|---|---|
| cocoa area | -0.019 | 0.013 | 0.014 | -0.014 | -0.147* | -0.147* | -0.644*** | -0.480*** |
| | (0.017) | (0.015) | (0.018) | (0.017) | (0.078) | (0.078) | (0.143) | (0.134) |
| treated cohort × cocoa area | 0.025 | 0.024 | -0.028 | -0.024 | 0.107 | 0.107 | -0.203 | -0.209 |
| | (0.025) | (0.021) | (0.028) | (0.026) | (0.100) | (0.100) | (0.168) | (0.161) |
| Constant | 0.171*** | 0.454*** | 0.719*** | 0.515*** | 5.639*** | 5.639*** | 7.433*** | 8.791*** |
| | (0.009) | (0.033) | (0.010) | (0.048) | (0.188) | (0.188) | (0.076) | (0.292) |
| Observations | 6,479 | 6,479 | 6,479 | 6,479 | 6,472 | 6,472 | 6,479 | 6,479 |
| R-squared | 0.076 | 0.222 | 0.053 | 0.111 | 0.097 | 0.097 | 0.067 | 0.126 |
| 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 | No | Yes | No | Yes |
| Outcome Mean | 0.165 | 0.165 | 0.723 | 0.723 | 5 | 5 | 7.103 | 7.103 |

The table displays the long-term effects of the swollen shoot cocoa disease labor market outcomes. The cocoa area refers to places climatically suited for cocoa cultivation and the treated cohort refers to all births between 1940–1950. All models include cohort (year of birth) fixed effects and region fixed effects. The controls are age of household head and indicator variables for female, rural, female head of household, and head of households. Standard errors are clustered at the cohort × region level. *** p<0.01, ** p<0.05, * p<0.1

Table 4: The Gender-Specific Effects of the Swollen Shoot Disease

| | (1) years of schooling | (2) formal education | (3) morbidity | (4) mortality | (5) waged employment | (6) self employment | (7) work days | (8) work hours |
|---|---|---|---|---|---|---|---|---|
| cocoa area | 0.948*** | 0.105*** | 0.006 | -0.017 | -0.016 | 0.013 | -0.180** | -0.629*** |
| | (0.209) | (0.018) | (0.018) | (0.011) | (0.016) | (0.017) | (0.079) | (0.144) |
| treated cohort × cocoa area | 0.168 | 0.025 | -0.006 | 0.037** | 0.086*** | -0.104*** | 0.087 | -0.201 |
| | (0.362) | (0.030) | (0.036) | (0.018) | (0.032) | (0.035) | (0.113) | (0.209) |
| female | -2.570*** | -0.196*** | 0.035*** | -0.012* | -0.211*** | 0.069*** | -0.404*** | -1.044*** |
| | (0.119) | (0.012) | (0.012) | (0.007) | (0.011) | (0.018) | (0.052) | (0.076) |
| treated cohort × cocoa area × female | -1.719*** | -0.163*** | 0.095** | -0.011 | -0.120*** | 0.140*** | 0.028 | -0.054 |
| | (0.393) | (0.041) | (0.042) | (0.020) | (0.030) | (0.040) | (0.133) | (0.209) |
| Constant | 5.881*** | 0.599*** | 0.288*** | 0.161*** | 0.284*** | 0.683*** | 5.294*** | 7.991*** |
| | (0.128) | (0.012) | (0.011) | (0.007) | (0.011) | (0.014) | (0.049) | (0.088) |
| Observations | 7,424 | 7,427 | 7,427 | 12,191 | 6,479 | 6,479 | 6,472 | 6,479 |
| R-squared | 0.331 | 0.342 | 0.036 | 0.040 | 0.165 | 0.064 | 0.094 | 0.099 |
| 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 | 4.805 | 0.531 | 0.315 | 0.150 | 0.165 | 0.723 | 5 | 7.103 |

The table displays the gender-specific long-term effects of the swollen shoot cocoa disease. The cocoa area refers to places climatically suited for cocoa cultivation and the treated cohort refers to all births between 1940–1950. All models include cohort (year of birth) fixed effects and region fixed effects. Model 4 also adds child's year of birth fixed effects. Standard errors are clustered at the cohort × region level. *** p<0.01, ** p<0.05, * p<0.1

Table 5: The Long-term Effects of the Swollen Shoot Disease Based on Type of Place of Residence

| | (1) years of schooling | (2) formal education | (3) morbidity | (4) mortality | (5) waged employment | (6) self employment | (7) work days | (8) work hours |
|---|---|---|---|---|---|---|---|---|
| cocoa area | 1.302*** | 0.131*** | 0.005 | -0.023** | 0.009 | -0.005 | -0.159** | -0.495*** |
| | (0.182) | (0.017) | (0.018) | (0.011) | (0.015) | (0.017) | (0.078) | (0.134) |
| treated cohort × cocoa area | -1.259** | -0.147*** | -0.022 | 0.024 | 0.068 | -0.069 | 0.260* | 0.082 |
| | (0.504) | (0.044) | (0.035) | (0.023) | (0.043) | (0.044) | (0.148) | (0.294) |
| rural | -2.690*** | -0.199*** | 0.004 | 0.042*** | -0.216*** | 0.145*** | -0.187*** | -1.109*** |
| | (0.157) | (0.015) | (0.015) | (0.008) | (0.013) | (0.015) | (0.062) | (0.100) |
| treated cohort × cocoa area × rural | 0.780 | 0.122** | 0.096*** | 0.008 | -0.052 | 0.052 | -0.193 | -0.356 |
| | (0.528) | (0.048) | (0.035) | (0.024) | (0.041) | (0.042) | (0.150) | (0.298) |
| Constant | 6.079*** | 0.610*** | 0.306*** | 0.128*** | 0.307*** | 0.629*** | 5.194*** | 8.127*** |
| | (0.138) | (0.012) | (0.012) | (0.008) | (0.013) | (0.015) | (0.060) | (0.104) |
| Observations | 7,424 | 7,427 | 7,427 | 12,191 | 6,479 | 6,479 | 6,472 | 6,479 |
| R-squared | 0.306 | 0.322 | 0.033 | 0.042 | 0.136 | 0.072 | 0.084 | 0.094 |
| 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 | 4.805 | 0.531 | 0.315 | 0.150 | 0.165 | 0.723 | 5 | 7.103 |

The table displays the long-term effects of the swollen shoot cocoa disease based on type of place of residence (rural/urban). The cocoa area refers to places climatically suited for cocoa cultivation and the treated cohort refers to all births between 1940–1950. All models include cohort (year of birth) fixed effects and region fixed effects. Model 4 also adds child's year of birth fixed effects. Standard errors are clustered at the cohort × region level. *** p<0.01, ** p<0.05, * p<0.1

Table 6: The Long-term Effects of the Swollen Shoot Disease Based on Householder Status

| | (1) years of schooling | (2) formal education | (3) morbidity | (4) mortality | (5) waged employment | (6) self employment | (7) work days | (8) work hours |
|---|---|---|---|---|---|---|---|---|
| cocoa area | 1.307*** | 0.130*** | -0.000 | -0.018 | -0.021 | 0.011 | -0.191** | -0.660*** |
| | (0.176) | (0.016) | (0.018) | (0.011) | (0.016) | (0.018) | (0.078) | (0.143) |
| treated cohort × cocoa area | -1.371*** | -0.101*** | 0.093** | 0.034** | -0.017 | 0.027 | 0.300** | -0.192 |
| | (0.357) | (0.038) | (0.040) | (0.018) | (0.028) | (0.046) | (0.150) | (0.230) |
| householder | -0.113 | 0.017 | 0.094*** | -0.002 | 0.138*** | 0.089*** | 0.177*** | 0.749*** |
| | (0.185) | (0.017) | (0.020) | (0.008) | (0.010) | (0.019) | (0.053) | (0.080) |
| treated cohort × cocoa area × householder | 0.874** | 0.054 | -0.060 | -0.007 | 0.058* | -0.076 | -0.254* | -0.008 |
| | (0.387) | (0.042) | (0.048) | (0.021) | (0.031) | (0.051) | (0.153) | (0.242) |
| Constant | 8.047*** | 0.741*** | 0.146*** | 0.156*** | 0.082*** | 0.662*** | 4.965*** | 6.949*** |
| | (0.368) | (0.037) | (0.039) | (0.006) | (0.010) | (0.015) | (0.058) | (0.089) |
| Observations | 7,424 | 7,427 | 7,427 | 12,191 | 6,479 | 6,479 | 6,472 | 6,479 |
| R-squared | 0.375 | 0.366 | 0.044 | 0.039 | 0.107 | 0.061 | 0.083 | 0.081 |
| 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 | 4.805 | 0.531 | 0.315 | 0.150 | 0.165 | 0.723 | 5 | 7.103 |

The table displays the long-term effects of the swollen shoot cocoa disease based on householder status. The cocoa area refers to places climatically suited for cocoa cultivation and the treated cohort refers to all births between 1940–1950. All models include cohort (year of birth) fixed effects and region fixed effects. Model 4 also adds child's year of birth fixed effects. Standard errors are clustered at the cohort × region level. *** p<0.01, ** p<0.05, * p<0.1

Table 7: The Long-term Effects of the Swollen Shoot Disease Based on Gender of Household Head

| | (1) years of schooling | (2) formal education | (3) morbidity | (4) mortality | (5) waged employment | (6) self employment | (7) work days | (8) work hours |
|---|---|---|---|---|---|---|---|---|
| cocoa area | 0.942*** | 0.104*** | 0.005 | -0.018 | -0.017 | 0.012 | -0.181** | -0.639*** |
| | (0.216) | (0.018) | (0.018) | (0.011) | (0.017) | (0.018) | (0.079) | (0.145) |
| treated cohort × cocoa area | -0.372 | -0.020 | 0.037 | 0.037** | 0.048* | -0.055* | 0.165 | -0.150 |
| | (0.293) | (0.026) | (0.027) | (0.018) | (0.027) | (0.030) | (0.108) | (0.182) |
| female head | -0.845*** | -0.046*** | 0.092*** | 0.004 | -0.104*** | 0.151*** | -0.266*** | -0.305*** |
| | (0.147) | (0.015) | (0.014) | (0.008) | (0.012) | (0.014) | (0.061) | (0.083) |
| treated cohort × cocoa area × female head | -1.285*** | -0.144*** | 0.036 | -0.013 | -0.080*** | 0.097*** | -0.181 | -0.190 |
| | (0.403) | (0.048) | (0.039) | (0.021) | (0.028) | (0.032) | (0.185) | (0.238) |
| Constant | 4.664*** | 0.501*** | 0.283*** | 0.154*** | 0.197*** | 0.681*** | 5.146*** | 7.511*** |
| | (0.117) | (0.010) | (0.010) | (0.006) | (0.010) | (0.011) | (0.046) | (0.081) |
| Observations | 7,424 | 7,427 | 7,427 | 12,191 | 6,479 | 6,479 | 6,472 | 6,479 |
| R-squared | 0.269 | 0.301 | 0.040 | 0.039 | 0.093 | 0.077 | 0.086 | 0.070 |
| 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 | 4.805 | 0.531 | 0.315 | 0.150 | 0.165 | 0.723 | 5 | 7.103 |

The table displays the long-term effects of the swollen shoot cocoa disease based on gender of household head. The cocoa area refers to places climatically suited for cocoa cultivation and the treated cohort refers to all births between 1940–1950. All models include cohort (year of birth) fixed effects and region fixed effects. Model 4 also adds child's year of birth fixed effects. Standard errors are clustered at the cohort × region level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

## 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, particularly in the forest zone, where climatic conditions are suitable for its cultivation. Due to its comparatively low costs, cocoa farming attracted many farmers who had previously engaged in other agricultural pursuits, such as oil palm and coffee cultivation.

However, beginning in the 1930s, the cocoa industry faced a significant threat: farmers began noticing a dieback of cocoa trees due to a virus, later known as the cocoa swollen shoot virus disease (CSSVD) as it caused distended branches. The disease would attack the tree in its entirety and eventually kill it. The virus spread alarmingly fast and caused devastating yield loss. By the mid-1940s more than 40 million trees had been infected, with the virus continuing to spread to 15 million trees annually. Farm output had also plummeted from 30 tons to 6 tons.

After various field experiments, it was determined that controlling the disease required ”roguing,” the careful detection and removal of infected trees. However, this method proved ineffective, as asymptomatic but infected trees remained. Consequently, authorities mandated the removal of both infected trees and their neighbors. As a result, cocoa farmers suffered not only from yield loss due to the virus but also from additional economic impacts due to mandatory tree removal. Furthermore, by the mid 20th century cocoa had become the most valuable export commodity in the Gold Coast, underscoring the potential severe impacts of the viral outbreak for cocoa-farming households.

In this study I investigate the long-term effects of this agricultural epidemic on the human capital and labor market outcomes of children born in cocoa-producing areas during the epidemic. The study is framed within the context of the fetal origins hypothesis which posits that early-life exposures, such as income shocks and prenatal stress/ruptures, can have lasting negative consequences. The CSSVD could have affected child development through several interrelated pathways, including reduced household income, limited parental investments in education and healthcare, maternal stress during pregnancy, and nutritional deficits due to lower food security.

Employing a cohort-based Difference-in-Difference (DID) approach, the findings reveal that the epidemic had a significant negative impact on human capital but no effect on labor market outcomes. Children born in cocoa-growing areas during the epidemic experienced three-quarters of a year loss in schooling and 6.3 percentage points (pp) decline in formal education. Their morbidity rate also increased by 5 pp, and they exhibited a 3 pp higher early mortality. Conducting a heterogeneous analysis, I find that these effects were mostly driven by girl children, urban populations, and female-headed households.

In contrast, children born in cocoa areas during the epidemic were neither less likely to be waged or self employed nor did they work fewer days or hours. This suggests that, unlike educational attainment and health – which are often predetermined by early-life conditions – labor market outcomes offer more flexibility for adaptation over time. Despite facing significant early-life disadvantages in terms of education and health, individuals may have been able to integrate into the labor market through avenues such as job training, work experience, or economic opportunities that arose later in life.

This study makes a significant contribution to the literature on early-life shocks and their long-term impacts on human capital and labor market outcomes. By focusing on an agriculture-related epidemic – a context that has received relatively little attention in the literature – this study provides new insights into how industry-specific shocks shape child development. The findings highlight the persistent effects of early-life economic hardship on education and health, while also revealing the potential for recovery in labor market outcomes. From a policy perspective, the results underscore the importance of timely support for vulnerable households during economic crises.

## 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.

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.

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.

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 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.

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

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.

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.

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: Using a Placebo "Treated Cohort"

| | (1)<br>years of schooling | (2)<br>formal education | (3)<br>morbidity | (4)<br>mortality | (5)<br>waged employment | (6)<br>waged employment | (7)<br>work days | (8)<br>work hours |
|---|---|---|---|---|---|---|---|---|
| cocoa area | 1.159*** | 0.120*** | 0.027 | -0.015 | 0.020 | -0.022 | -0.153* | -0.497*** |
| | (0.189) | (0.017) | (0.018) | (0.012) | (0.014) | (0.017) | (0.080) | (0.133) |
| treated cohort × cocoa area | -0.030 | -0.011 | -0.065*** | -0.001 | -0.007 | 0.010 | 0.097 | -0.092 |
| | (0.222) | (0.021) | (0.022) | (0.013) | (0.019) | (0.022) | (0.091) | (0.166) |
| Constant | 8.024*** | 0.741*** | 0.151*** | 0.127*** | 0.453*** | 0.516*** | 5.630*** | 8.803*** |
| | (0.369) | (0.037) | (0.039) | (0.021) | (0.033) | (0.048) | (0.187) | (0.292) |
| Observations | 7,424 | 7,427 | 7,427 | 12,191 | 6,479 | 6,479 | 6,472 | 6,479 |
| R-squared | 0.373 | 0.365 | 0.042 | 0.047 | 0.222 | 0.111 | 0.097 | 0.126 |
| 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 | 4.805 | 0.531 | 0.315 | 0.150 | 0.165 | 0.723 | 5 | 7.103 |

The table displays the placebo effects of the cocoa swollen shoot virus disease in the Gold Coast. The cocoa area refers to places climatically suited for cocoa cultivation and the "treated cohort" refers to all births between 1951–1960. All models include cohort (year of birth) fixed effects and region fixed effects. The controls are age of household head and indicators for female, rural, female head of household, and head of households. Standard errors are clustered at the cohort × region level. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table A2: Robustness Check: Using Placebo "Treated Regions"

| | (1)<br>years of schooling | (2)<br>formal education | (3)<br>morbidity | (4)<br>mortality | (5)<br>waged employment | (6)<br>self employment | (7)<br>work days | (8)<br>work hours |
|---|---|---|---|---|---|---|---|---|
| non-cocoa regions | -2.031*** | -0.241*** | -0.025* | 0.070*** | -0.020 | -0.055*** | 0.730*** | 0.504*** |
| | (0.247) | (0.026) | (0.013) | (0.008) | (0.012) | (0.014) | (0.068) | (0.109) |
| treated cohort × non-cocoa regions | 0.592 | 0.039 | -0.054* | -0.022 | -0.015 | 0.009 | -0.085 | 0.297 |
| | (0.471) | (0.046) | (0.028) | (0.0.016) | (0.027) | (0.029) | (0.156) | (0.217) |
| Constant | 10.063*** | 0.965*** | 0.176*** | 0.081*** | 0.508*** | 0.525*** | 5.181*** | 8.394*** |
| | (0.395) | (0.039) | (0.039) | (0.021) | (0.032) | (0.049) | (0.188) | (0.288) |
| Observations | 7,424 | 7,427 | 7,427 | 12,191 | 6,479 | 6,479 | 6,472 | 6,479 |
| R-squared | 0.317 | 0.299 | 0.033 | 0.036 | 0.207 | 0.105 | 0.068 | 0.104 |
| Cohort F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Region F.E. | No | No | No | No | No | No | No | No |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Outcome Mean | 4.805 | 0.531 | 0.315 | 0.150 | 0.165 | 0.723 | 5 | 7.103 |

The table displays the placebo effects of the swollen shoot cocoa disease in the Gold Coast. The non-cocoa regions refers to places climatically unsuited for cocoa cultivation and the treated cohort refers to all births between 1940–1950. All models include cohort (year of birth) fixed effects and region fixed effects. The controls are age of household head and indicators for female, rural, female head of household, and head of households. Standard errors are clustered at the cohort × region level. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table A3: Robustness Checks: Effects of Migration Between Cocoa and Non-cocoa Regions

| | (1) years of schooling | (2) formal education | (3) morbidity | (4) waged employment | (5) self employment | (6) work days | (7) work hours |
|---|---|---|---|---|---|---|---|
| cocoa area | 1.249*** | 0.123*** | 0.002 | 0.014 | -0.021 | -0.140* | -0.489*** |
| | (0.175) | (0.016) | (0.018) | (0.015) | (0.017) | (0.080) | (0.136) |
| treated cohort × cocoa area | -0.848*** | -0.079*** | 0.036 | 0.025 | -0.026 | 0.125 | -0.103 |
| | (0.267) | (0.025) | (0.027) | (0.021) | (0.026) | (0.110) | (0.170) |
| cocoa to non-cocoa | -0.175 | 0.009 | 0.056*** | -0.042* | 0.067** | -0.015 | -0.239* |
| | (0.220) | (0.022) | (0.020) | (0.021) | (0.026) | (0.087) | (0.137) |
| non-cocoa to cocoa | -0.917*** | -0.113*** | -0.015 | 0.016 | -0.079*** | 0.102 | 0.054 |
| | (0.246) | (0.023) | (0.020) | (0.018) | (0.022) | (0.085) | (0.162) |
| cocoa to non-cocoa × treated cohort | -0.377 | -0.072** | -0.057 | 0.012 | -0.006 | 0.107 | 0.607*** |
| | (0.449) | (0.035) | (0.041) | (0.033) | (0.044) | (0.171) | (0.207) |
| non-cocoa to cocoa × treated cohort | 0.532 | 0.046 | 0.034 | 0.022 | 0.033 | 0.052 | 0.356 |
| | (0.473) | (0.049) | (0.047) | (0.037) | (0.044) | (0.203) | (0.306) |
| Constant | 8.115*** | 0.749*** | 0.142*** | 0.456*** | 0.513*** | 5.626*** | 8.790*** |
| | (0.368) | (0.037) | (0.039) | (0.033) | (0.048) | (0.188) | (0.292) |
| Observations | 7,424 | 7,427 | 7,427 | 6,479 | 6,479 | 6,472 | 6,479 |
| R-squared | 0.376 | 0.369 | 0.044 | 0.223 | 0.115 | 0.097 | 0.127 |
| Cohort F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Region F.E. | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Outcome Mean | 4.805 | 0.531 | 0.315 | 0.165 | 0.723 | 5 | 7.103 |

The table displays the long-term effects of the swollen shoot cocoa disease accounting for the possible effects of migration between cocoa and non-cocoa regions. The cocoa area refers to places climatically suited for cocoa cultivation and the treated cohort refers to all births between 1940–1950. "Cocoa to non-cocoa" and "Non-cocoa to cocoa" are indicator variables for migrants moving from cocoa to non-cocoa regions and from non-cocoa regions to cocoa regions respectively. All models include cohort (year of birth) fixed effects and region fixed effects. The controls are age of head of household and indicator variables for female, rural, female head of household, and head of household. Standard errors are clustered at the cohort × region level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$