Orphans and Schooling in Africa: A Longitudinal Analysis

David Evans*
Harvard University

Edward Miguel
University of California, Berkeley and NBER

July 2004

Abstract: AIDS deaths could have a long-run impact on economic development by affecting the human capital accumulation of the next generation. We estimate the impact of parent death on primary school participation using an unusual five-year panel data set of over 18,000 children in rural Kenya. We restrict attention to children who began the study period as non-orphans, and compare those who subsequently lost a parent to those who did not. Parent death leads to a moderate decrease in school participation, with larger effects for children from poor households. Impacts are not significantly different by child gender, nor do they significantly depend on local orphan rates. We discuss implications for economic growth in Africa, and for the design of programs to assist AIDS orphans.

JEL Classification: I12, I20, J13, O12

---

* **Acknowledgements:** We thank ICS Africa and the Kenya Ministry of Health Division of Vector Borne Diseases for their cooperation in all stages of the project, and would especially like to acknowledge the contributions of Elizabeth Beasley, Laban Benaya, Pascaline Dupas, Simon Brooker, Alfred Luoba, Sylvie Moulin, Robert Namunyu, Polycarp Waswa, and the PSDP field staff and data group, without whom the project would not have been possible. Harold Alderman, Gustavo Bobonis, Larry Katz, Michael Kremer, and participants in seminars at Harvard University and the World Bank have provided valuable comments. Melissa Gonzalez-Brenes and Tina Green have provided excellent research assistance. We are grateful for financial support from the World Bank, the Partnership for Child Development, the Harvard Center for International Development, the Harvard University Inequality Program, and the National Institutes of Health Fogarty International Center. All errors are our own.

D. Evans, Department of Economics, Harvard University, dkevans@fas.harvard.edu.
E. Miguel, Department of Economics, University of California, 549 Evans Hall #3880, Berkeley, CA 94720-3880, USA , emiguel@econ.berkeley.edu, phone 1-510-642-7162, fax 1-510-642-6615.

1. Introduction

More than one in nine children under 15 years of age in Sub-Saharan Africa in 2001 had lost a parent, and the HIV/AIDS pandemic is the leading cause (UNAIDS 2002).[2] HIV/AIDS deaths today could potentially have long-run effects on economic development by affecting the human capital accumulation of the next generation. Yet although children orphaned by AIDS have received considerable international media coverage,[3] little systematic empirical research has estimated the impact of parent death on these children. Understanding the magnitude of these effects – both on average, and for households and communities with particular characteristics – is critical for the design of programs to successfully assist orphans.

This paper uses an unusual panel data set gathered over nearly five years to measure the impact of parent mortality on school participation in rural western Kenya. The data set was originally collected to evaluate a school health project. This study has several strengths. First, we are able to estimate longer-term impacts of parent death than existing studies, as well as effects in the years preceding the parent death – due to AIDS-related morbidity, for instance. We restrict attention to children who began the study period as non-orphans, and compare school participation of those who subsequently lost a parent to those who did not in order to partially address omitted variable bias concerns.

Second, rich baseline data on child and household characteristics allow us to control for any differences across children before parent death, and to test if parent death has differential impacts for children from particular types of households. Finally, the sample contains communities with a wide range of orphan rates, allowing us to estimate how the local context affects orphan outcomes, and in particular whether community safety nets break down in areas with many orphans.

[2] In this paper, as in most of the literature on orphanhood in Sub-Saharan Africa, a child is referred to as an "orphan" if her mother has died, if her father has died, or both.

[3] For two examples, see Robinson (1999) and Wax (2003).

It is worth stressing up front that we neither have individual biomedical information on HIV infection status, nor data on whether the cause of a parent death was AIDS: while HIV prevalence in the study region is estimated to lie between 20 to 30 percent (NASCOP 2001), and thus it is likely that many adult deaths are HIV/AIDS-related, we cannot determine the exact proportion in our sample. Since we lack data on the cause of parent death, we are unable to test whether AIDS orphans fare differently than other orphans due to any stigma attached to AIDS, for instance. Nonetheless, UNAIDS et al. (2002) estimate that in Kenya as a whole, where the HIV prevalence rate was estimated at 15 percent in 2001, 54 percent of orphans up to 14 years old had lost at least one parent to AIDS – and this proportion is likely to be even greater in our study region (Busia district), where prevalence is considerably higher. Thus we feel confident in asserting that the bulk of the orphans in our sample are AIDS orphans.

There is a moderate and highly statistically significant negative impact of a parent death on primary school participation: school participation rates declined on average by 3 percentage points following a parent death, and impacts are nearly identical for deaths of mothers versus fathers. Contrary to some observers' claims, child age and gender are not associated with differential parent death impacts on school participation. The proportion of orphan children in the community similarly does not differentially affect school participation after a parent death, providing some evidence against the view that family and community networks are rapidly disintegrating in African areas hardest-hit by HIV/AIDS – although limited statistical precision means this last result should be interpreted with caution.

In a further result, the negative impact of parent death is concentrated in the households with fewer assets before the parent death: the drop in school participation is 9 percentage points for households with no latrine or toilet at baseline (these households are roughly one quarter of the sample). One policy implication of this finding is that assistance programs should target their efforts to poor orphans in rural African settings. Unfortunately, the data set does not contain detailed information on household income or consumption expenditures, so we cannot directly test whether the estimated school participation impacts are caused by households' inability to pay primary school fees, or through other channels – such as psychological trauma, or the lack of parental encouragement and emotional support[4] – although the finding that effects are larger for children from poor households suggests that income is playing a role.

We attempt to rule out the possibility that the estimated relationships are spurious – driven by unobserved differences between households where parents die versus households where they survive – in several ways. First, the results are robust to numerous baseline household and community characteristics, and to individual fixed effects. Second, a range of baseline (pre-death) characteristics are nearly identical for the children who experience a parent death during the period and for those who do not, and pre-death school participation levels and trends are also similar. While not absolutely definitive, this paper's econometric approach constitutes a considerable improvement over existing studies.

Yamano and Jayne (2003) is the most closely related work to the current study. They use a two-year panel data set of Kenyan household surveys, and a difference-in-differences identification strategy related to the one we employ in this paper, to estimate the impact of parent death on school enrollment. Yamano and Jayne (2003) find significant negative impacts of parent death on school enrollment, but only among poor children, echoing our results. Yamano and Jayne (2003) also find effects prior to the parent death, as we do in this study, although our larger sample size, longer time period, and continuous data collection throughout the study period (the Yamano and Jayne data set only has two observations, for 2000 and 2002) allows us to more precisely measure parent morbidity and death impacts. Another advantage of our approach is the use of school attendance and enrollment data collected at school by enumerators during unannounced visits, rather than relying on parent (or guardian) reports of school enrollment (Yamano and Jayne 2003 rely on the latter).[5] Gertler et al. (2004) estimate similar parent

[4] Gertler et al. (2003) suggest that the negative impact of parent death on schooling in Indonesia is almost entirely explained by causes other than the loss of income, but this may or may not also hold in Africa.

[5] In a related paper, Yamano and Jayne (2004) find large and statistically significant negative impacts of the death of male household heads on the net value of household crop production – a massive drop of 68 percent – as well as in off-farm income, and speculate that an income shock of this magnitude would negatively impact the school enrollment of children in the affected household, by making it difficult to pay school fees. For the death of a female

death impacts in Indonesia during the mid-1990s: a parent death during the previous twelve months leads to a doubling in the probability a child drops out of school that year. However, an important limitation is their inability to estimate impacts of parent death over periods longer than one year.

The remainder of the paper is structured as follows. Section 2 describes the data set and discusses sample attrition, section 3 outlines the estimation strategy, section 4 presents the main empirical results and derives some macroeconomic implications, and the final section concludes.

2. Data and Measurement

The data set was collected in Busia district, Kenya – a densely settled farming region adjacent to Lake Victoria – in the context of the evaluation of a primary school health program, which provided medical treatment for intestinal helminth (worm) infections. The Kenyan non-governmental organization (NGO) ICS Africa began carrying out that program in late 1997, and the 75 schools taking part consist of nearly all rural primary schools in Budalangi division and Funyula division in southern Busia. (The program is described in Miguel and Kremer 2004.)

The foremost strength of this education data set, and what sets it apart from most other African data sets, is its length of over four years, from early 1998 to mid-2002. A second strength is that schools were visited multiple times each year by enumerators to record student school attendance and enrollment, where these visits were not announced to the school in advance. This results in a more compelling measure of school participation than either data collected from primary school attendance registers, which are thought to be unreliable in less developed countries, or parent responses about the school enrollment status of their children. School participation varies with the agricultural season and with household economic needs in many poor countries, and thus parent reports on current enrollment status may convey little information about actual school attendance patterns.

---

household head, or of the female spouse of the male household head, the measured impact on crop production is also large, though somewhat smaller (at 46 percent) and not statistically significant.

The second dataset we use is the 2002 Tracking Survey. The original 75 schools were visited by enumerators between February and August 2002, in order to track each child from the 1998 baseline sample. If the child was present at school, she was asked directly about the mortality of her parents, and if a parent had died, about the exact year of death. If the child was not present at school that day, teachers and other students in the school were asked to provide this information. In practice, it was common for siblings, cousins, and neighbors of absent children to volunteer the information on parent death, as this information appears to be widely known in rural Kenyan communities. As a check on data reliability, the parent mortality data collected at school was compared to mortality data collected at children's homes in 2001 for a representative sub-sample of 69 of the children (these home survey data were typically collected from intimate relatives of the dead parent). There is a high correlation in the reported year of father death between the two surveys (correlation coefficient 0.87), but the correlation for mother deaths is considerably lower (at 0.61) – although since there are many fewer mother deaths than father deaths, this latter figure is based on relatively few observations.

It is worth noting that we are likely to underestimate parent death impacts to the extent that this indirect survey method in the 2002 Tracking Survey captures information on the exact year of parent death with some error, under plausible assumptions on measurement error (Aigner 1973). Another reason we may underestimate true effects is the possibility of parent morbidity in the year (or years) before their death; in this case, the difference between child school enrollment immediately before and after the parent death understates the total effect of parent illness and death taken together – although in our case, we are also able to estimate pre-death morbidity effects, using methods discussed below. A third reason that true effects may be underestimated is that this data set does not include information on future parent deaths. In other words, in 2002 data, children who (unbeknownst to the econometrician) will experience a parent death in, say, 2003 or 2004 could already be experiencing adverse impacts due to AIDS-related morbidity, and this is likely to reduce the estimated impact of parent illness for children whose parents had already died by 2002.

The third data set, the 1998 Pupil Questionnaire, was administered in January to March 1998 and collected information from children on a variety of health measures and household socioeconomic characteristics, providing valuable baseline (pre-parent death) controls for a subset of children. Finally, we employ the 2002 Headmaster Questionnaire, which collected information from primary school headmasters regarding policies toward orphans in their school in a representative subsample of 48 of the 75 program schools during May to July 2002.

2.1 Sample Size and Attrition

The existence of very few panel data sets in Sub-Saharan Africa is in part due to difficulties in tracking survey respondents through time, and we are not immune to this problem. Household migration, child fostering, and imperfect recall in the 2002 Tracking survey all complicated our task, and led to non-trivial rates of both missing data and sample attrition.

There are two sources of missing observations that reduce our "baseline sample" of 24,111 children. This baseline sample includes all children who were not orphans at baseline in early 1998, were enrolled in the 75 NGO program schools in grades 1 through 7, and were between 5 and 18 years old. We use two samples of children in the analysis, the "full sample" and the "restricted sample." The full sample of 18,133 children includes all baseline students for whom there is parent mortality data.[6] Most cases of unknown or unreliable orphan status were among children initially in the upper grades in 1998, who had been out of school longer than younger pupils (and thus were often not as well-known to other children in the school). These individuals without reliable parent death information tend to have worse school participation outcomes than other students, so to the extent that parent death information is more likely to

[6] Age data are missing for 3,163 children in the full sample; it is common for individuals not to know their birth date or year in rural Kenya, where few individuals obtain formal birth certificates. However, we include these observations in the analysis using indicator variable controls that take on a value of one for observations with missing values. Individuals are dropped from the sample when they reach 18 years of age, due to the difficulty in collecting reliable schooling information for them past that age.

be missing for orphans (who may have been more likely to move to distant geographic districts after the death) than for non-orphans, our estimates would understate true parent death effects.

The restricted sample contains 7,815 children from the full sample for whom baseline individual and household measures from the 1998 Pupil Questionnaire data set are available. The restricted sample first drops all 6,718 students (of the 18,133 students in the full sample) initially enrolled in grades 1 and 2 in 1998, since there is no detailed pupil survey information on them (the 1998 pupil survey was only administered in grades 3 and higher). Of the remaining 11,415 students, there is pupil survey information for only a subset of students in grades 3 to 7, those who were present at school on the day of the survey, leaving 7,815 children. Finally, of the 7,815 students in the restricted sample, 2,194 attrit at some point during the study; that is, there is school participation data for them for only a subset of the four years we study. In many cases, these students moved away from the area and there is no information regarding their subsequent school enrollment. These children are included in the analysis only in years that we observe them.

Child characteristics are similar for both the full and restricted samples (Table 1). A relatively large proportion (8 percent) of children who were non-orphans in 1998 became orphans during 1999-2002. Fully 15 percent of schoolchildren in this area were orphans at baseline in 1998, and this proportion varies widely across the 75 communities, from near zero in some areas up to 40 percent in others, and this variation allows us to estimate the impact of parent death across communities with very different rates. Children in this region are quite poor even by Kenyan standards, and this translates into poor health and nutrition (Table 1, Panel B): nearly 20 percent of households lack a latrine (or toilet) at home, only 14 percent of children wore shoes to school in 1998, and almost two-fifths experienced a fever in the month previous to the survey.[7] The average weight-for-age z-score is -1.44, which is similar to the overall average for Kenyan children (UNDP 2002).

We next test whether those students with missing data are significantly different from other students along observable dimensions, specifically in terms of age, gender, geographic location of the school, and in some specifications household socioeconomic characteristics. In the first case, the dependent variable is an indicator that takes on a value of one if the child is missing parent death information.[8] In the baseline sample of 24,111 children, the older children, those in Budalangi division, and unsurprisingly those with missing age data are much more likely to have missing orphan status information (Table 2, regression 1). Among the subsample with detailed baseline characteristics from the 1998 pupil survey, there is no clear effect of socioeconomic status on missing orphan status information (regression 2), although certain characteristics are statistically significant; for instance, children wearing shoes in the baseline survey were somewhat more likely to have missing data, while the opposite holds for children wearing a uniform.

Considering attrition (in other words, missing school participation) as the dependent variable, children who we know become orphans during the period are significantly *less* likely to attrit among the 18,133 children for whom parent mortality data is available (Table 2, regressions 3 and 4).[9] This result appears counter-intuitive at first, but is consistent with the notion that orphan status is more likely to be missing altogether for children who have attrited. However, if even a small fraction of the children with unknown orphan status are in fact orphans, then the coefficient estimate on the orphan indicator would switch signs and become positive; for instance, if we assign a random subset of 12 percent of the children whose orphan status is unknown (much less than the actual proportion of orphans in this population – see

[7] In practice, this includes disease episodes classified by children as either "fever" or "malaria" in the survey. Although poor children were asked to report recent cases of malaria (as opposed to fevers), fevers of any cause are often reported as malaria in areas where testing is costly (Watson 1992).

[8] This dependent variable has a mean of 0.25, i.e., 25 percent of the baseline sample is missing parent death data.

[9] We find that 28 percent of children attrit at some point during the study.

Table 1) to be “orphans” in a simulation, the coefficient estimate on the orphan term in regressions 3 and 4 becomes positive (results not shown). Results are similar when examining attrition in either the full sample or the restricted sample.

2.2 Primary School Fees and Orphans

The primary school finance context in rural Kenya is also important in understanding households’ school participation decisions.[10] Both the central government and local school committees play important roles in Kenyan primary school finance. The national Ministry of Education pays teacher salaries, while school committees raise funds locally for books, chalk, classrooms, and desks. Although the teacher salaries and benefits paid by the central government account for most primary school spending – approximately 90 percent (see Miguel and Gugerty 2002) – a reduction in local funding could have an important impact on educational outcomes if local inputs and teachers are complements in educational production. Local funds also have an important effect on public perceptions of school quality because they are largely used for construction and other visible capital projects.

Parents raise the bulk of local school funds through two mechanisms: school fees and village fundraising events (called *harambees* in Swahili). Annual school fees are set by the school committee – each primary school is managed by its own committee – and collected by the headmaster. School fees ranged roughly from 4-10 U.S. dollars per family during 1998-2002, a non-trivial amount in this area. The other important source of local primary school funding in western Kenya, accounting for approximately one-third of total local funding, are the village fundraisers. At these events parents and

[10] This section describes Kenyan primary school finance before Mwai Kibaki was elected president in December 2002. Starting in early 2003, the Kenya Ministry of Education abolished local primary school fees nation-wide, and agreed to provide some additional resources to primary schools to compensate for lost local funds. In future work, it would be interesting to estimate the impact of this local funding change on the school participation of orphans; for instance, Deininger et al. (2003) find that the introduction of a universal primary education reform in Uganda that abolished fees negated any adverse school enrollment impacts for foster children.

other community members meet and publicly pledge financial support for a planned school investment project, such as the construction of a new classroom. While contributions at these events are theoretically voluntary, school committees often set expected *harambee* contribution levels for parents and teachers in practice, and contributions are sometimes recorded by the school committee.

A variety of informal sanctions can be employed against parents if they fail to make school fee and *harambee* payments, including publicly naming them at community meetings, and temporarily suspending their children from school. While the threat of such sanctions is an important tool used by headmasters and school committees to enforce payment, the children of non-contributing parents cannot permanently be removed from school. Former President Moi of Kenya repeatedly stated that no child could be refused a primary education because of nonpayment of fees, and while official Ministry of Education policy was unclear, in practice this decree limited the discretion of schools to expel students. The content of threatened sanctions therefore does not include complete exclusion from public education, only temporary suspension de facto.

Few primary schools in this area make special allowances for orphans in terms of school fee reductions, according to the 2002 Headmaster Questionnaire. Forty-two of the 48 headmasters stated that orphans were subject to exactly the same school fees as other children. Of the 38 headmasters admitting they had sent some students temporarily away for nonpayment of school fees in the previous year, 32 claimed that orphans had been sent away just as often as non-orphans. More than two-thirds of the headmasters suggested that a main cause of dropping out for orphans was nonpayment of school fees, or not owning a school uniform (another large financial cost of schooling). Thus the inability to pay school fees is a plausible cause for drops in school participation after parent death to the extent that parent death reduces household disposable income (as in Yamano and Jayne 2004).

3. Empirical strategy

3.1 Identification Issues in Existing Research

Several recent studies have examined the issue of parent death and child schooling in less developed countries using a variety of methods and data sources, and they often yield quite different results. Most existing studies estimate observed differences between orphans and non-orphans at a single point in time, controlling for a limited set of current observable child characteristics. The results of such studies may be misleading due to both omitted variables and endogeneity: in the absence of longitudinal data, it is impossible to know whether these orphans and non-orphans were comparable before the parent death, and most importantly, current child and household characteristics may have themselves been affected by the death. Moreover, since parent death is relatively rare in most populations, few studies have sufficient statistical precision to reliably estimate small impacts.

Case et al. (2004) employ Demographic and Health Surveys (DHS) collected across ten Sub-Saharan African countries between 1992 and 2000 to estimate the impact of parent death on school enrollment. The large number of survey rounds (19 in all), combined with the relatively high incidence of parent death in their Africa sample, allows them to precisely estimate the impact of parent death. Their main finding is that orphans are significantly less likely to be enrolled in school than non-orphans. However, despite their impressive data effort, the study is likely to suffer from the omitted variable bias problem mentioned above, since they use cross-sections in the analysis rather than a panel. [11]

Earlier studies do not estimate negative parent death impacts on child education. Ainsworth et al. (2002) analyze a panel of 1,213 children in northwestern Tanzania, and find minimal impacts of parent death on schooling. In particular, child school enrollment is unaffected by parent death for non-poor households, while for poor households, they find that enrollment is merely delayed for the youngest children, but basically unaffected for older children. Note that, although Ainsworth et al. (2002) control

[11] The DHS household asset information is collected contemporaneously with the measurement of child orphan status, and thus is potentially endogenous: households fostering orphans may choose to sell off some assets, becoming poorer. It would be methodologically preferable to measure household characteristics prior to the parent death. The data used in Case et al. (2004) also give no indication of how long a child has been an orphan, and thus cannot shed light on how orphans fare through time before and after the parent death.

for baseline household characteristics, they do not use child fixed effects. Several studies echo Ainsworth et al. (2002) in finding little or no difference between orphans and non-orphans in terms of school enrollment (see Kamali et al. 1996, Ryder et al. 1994, Lloyd and Blanc 1996), although these all rely on less conclusive cross-sectional analysis. A number of international organization reports have claimed, however, that there are gender differences in parent death impacts on schooling, with girls suffering more than boys (World Bank 2002, UNAIDS 2002).

The absence of consistent negative estimated impacts of parent death on children in Sub-Saharan Africa in existing work has sometimes been attributed to the strength of extended family and community networks that care for orphans (Foster et al 1995, Foster and Williamson 2000, Ntozi 1997). One explanation for differences across settings, then, is the possibility that these insurance networks weaken or break down when local orphan rates surpass a certain critical level. However, the large estimated effects in Indonesia (Gertler et al., 2004) and small effects in Tanzania (Ainsworth et al. 2002) do not seem to fit this interpretation, given the much higher rate of orphanhood in Tanzania.

An alternative explanation for the small estimated orphan effects in cross-sectional studies is the possibility that HIV/AIDS victims are of somewhat higher socioeconomic status than non-victims, at least at the start of the epidemic; this will be the case if individuals in certain occupations particularly vulnerable to infection – including truckers, soldiers, and teachers – tend to be relatively affluent. To the extent that socio-economic variation is at least partially unobserved by the econometrician, this would lead to an upward bias in the estimated "impact" of being an orphan on subsequent life outcomes in cross-sectional studies, and may obscure negative parent death impacts.

3.2 Estimation Approach

The main estimation method in the current paper is linear regression with child fixed effects, where the "events" of interest are parent deaths. The child fixed effect captures time-invariant child and household characteristics that affect school participation. We compare changes in school participation (a variable which takes on values between zero and one) over time for children whose parents died during the period

1998-2002, to changes for children whose parents did not die. In some specifications we examine effects on school enrollment – an indicator variable that takes a value of one for students who were present in school during at least one enumerator visit during the course of the school year – as an alternative schooling outcome. The basic regression specification is as follows (where the "1" subscript refers to the equation number):

$$Y_{ijt} = \alpha_{1ij} + \rho_{1jt} + \beta_1 \cdot ORPHAN_{ijt} + \sum_c \gamma_1^C \cdot 1(c = C)_{ijt} + \delta_1 T_{jt} + u_{1i} + e_{1ijt} \qquad (1)$$

Y_{ijt} is the school participation rate for student *i* in school *j* during year *t*, α_{1ij} is a student fixed effect, ρ_{1jt} is a region-year indicator variable (at the level of the administrative division), and $ORPHAN_{ijt}$ takes a value of one if the individual is an orphan in period *t* (in other words, for all years during and after the parent death), and zero otherwise. In order to account for cohort and year-specific trends and gender differences in school participation which are independent of parent death, a full set of age cohort-year-gender indicator variables (where *c* denotes a particular age cohort-gender group in a particular year, e.g., girls who were 12 years old in 1998, observed in 1999) are included, $\sum_c \gamma_2^C \cdot 1(c = C)_{ijt}$. We also include an indicator variable for medical treatment through the school-based deworming program in school *j* in year *t*, T_{jt}, which was found to be related to school participation in existing work (Miguel and Kremer 2004). The coefficient β_1 captures the effect of parent death on school participation. To the extent that the unobserved differences between children who become orphans and those who do not are time-invariant, then equation 1 yields unbiased estimates of the effect of parent death on child outcomes (we discuss below what happens when we relax this assumption). In some specifications, baseline household socioeconomic and demographic characteristics are included as explanatory variables, rather than child fixed effects, and this generates broadly similar results, as discussed below.

There are a number of limitations to this approach. The specification imposes a constant effect of parent death on subsequent child outcomes regardless of when the parent died. It is more plausible that the effects of parent death might either compound over time (i.e., children whose parents died long ago experience increasingly adverse outcomes) or perhaps diminish if there are adequate coping mechanisms. To allow for such effects, we include indicator variables $\sum_{\tau} \beta^{S} 1(\tau = S)_{ijt}$, where τ is the number of years since the parent death; τ also takes on negative values in years before the parent death, for instance due to AIDS-related morbidity. Medical researchers estimate that AIDS deaths in nearby rural Uganda are typically preceded by 4 to 17 months of AIDS-related illness (Morgan et al. 2000, Morgan and Whitworth 2001), and thus we might expect negative effects up to two years before the parent death. In practice, we include indicators for each year from three years before the parent death to three years after the death (where the omitted category is four years before the parent death, or no parent death during this period). We do not observe children four or more years after a parent death, since the sample is restricted to children who were non-orphans at baseline in 1998 and we only observe the children through 2002. This leads to the preferred specification in equation 2:

$$Y_{ijt} = \alpha_{2ij} + \rho_{2jt} + \sum_{\tau} \beta_{2}^{S} \cdot 1(\tau = S)_{ijt} + \sum_{c} \gamma_{2}^{C} \cdot 1(c = C)_{ijt} + \delta_{2} T_{jt} + u_{2i} + e_{2ijt} \qquad (2)$$

Parent death may also have differential effects based on the gender of the parent; for instance, to the extent that mother's income and care-giving are more important than father's income and care-giving, mother death will have a greater impact on child outcomes (or vice versa). To estimate differential effects based on the gender of the parent, we include separate indicators for maternal death and paternal death, and also explore the impact of the first parent death versus the second parent death on child outcomes in some specifications. Finally, the parent death indicators are also interacted with household and community characteristics to test whether individuals from particular types of households or communities are differentially affected by parent death. For example, the magnitude of the parent death effect may depend on child age since older children are better substitutes for parents in the labor market, perhaps making them more likely to drop out of school after an adverse household income shock. We also test whether orphans fare worse in areas where there have already been more parent deaths, presumably due to a breakdown in local safety nets.

The key limitation of this econometric identification strategy is the possibility of time-varying family-level shocks affecting both parent health and child schooling outcomes. The most plausible sources are local weather or crop price shocks, but these are captured in the region-year indicator variables (ρ_{jt}) included in all main specifications. Another example of such a shock is parent job loss; however, in this rural area few individuals have formal sector jobs to lose. Child morbidity due to HIV infections contracted through parents is unlikely to affect the estimation, since the overwhelming majority of children born HIV-positive in rural Africa die before reaching school age (Adetunji 2000).[12]

We restrict attention to individuals whose parents were both alive at baseline in January 1998, and compare individuals whose parents subsequently died during the period 1999-2002 to those whose parents did not die, and we make the case that that these two groups of individuals – the "Became orphan" and "Never orphan" groups – are comparable along a range of observable characteristics at baseline. There are no significant differences in terms of baseline school participation or demographic characteristics in the full sample (Table 3, Panel A). In the restricted sample, the two groups look remarkably similar along twelve baseline characteristics, including measures of child nutrition and health, and household socioeconomic status (Panel B). There are small and marginally significant differences in child cleanliness and age, but we cannot reject that this pattern of differences is due to chance at traditional confidence levels. Further, for the 2,923 students for whom we have 1997 school participation data – gathered as part of another education intervention in a subset of schools – there is no significant difference between the 1997-1998 trends in school participation for children who later become orphans and those who do not (Panel C) – further evidence that they were in fact similar with respect to schooling before parent death.

These arguments do not completely eliminate concerns about the suitability of the comparison group, and in the absence of a clean quasi-experiment, it may be impossible to. Yet we feel that we are

[12] Moreover, child mortality rates are not statistically significantly different among those who become orphans and those who not, although note that there are few child deaths during the sample period (regressions not shown).

able to allay most potential concerns about the comparability of the two groups. If “Became orphan” and “Never orphan” households indeed differed sharply along unobserved dimensions – for instance, parents’ commitment to their children’s education, or their intertemporal discount rate – it is likely that these differences would also be reflected along some observable dimension given the rich set of characteristics we employ, but we do not find such differences.

Finally, note that children with unknown orphan status differ from other children in terms of lower baseline school participation and less household asset ownership (Table 3). To the extent orphans are more likely to drop out of school and leave the area, it is likely that these children are also disproportionately orphans, in which case excluding them from the analysis may lead us to underestimate actual parent death impacts.

4. Empirical Results

4.1 Parent Death Impacts

Parent death has a moderate negative impact on child school participation, defined as the total proportion of unannounced school attendance checks in which the child is present: on average, school participation falls by 0.037 (standard error 0.012) in a specification including individual baseline heath and socioeconomic controls (Table 4, regression 1), and is similar but somewhat smaller, at 0.026 (standard error 0.014), when individual fixed effects are included to reduce omitted variable bias (regression 2). All specifications include age cohort-year-gender indicator variables to capture differential enrollment patterns by demographic group, as well as region-year controls.[13]

While school participation is similar three years before parent death, it drops sharply in the year of the death and remains at the lower level for at least three years afterwards (Table 4, regressions 3 and 5). In other words, there is no evidence of orphan recovery after parent death. Figure 1 graphically

[13] The analogous estimate in Yamano and Jayne (2003) for a specification with child and household controls is somewhat larger, at -0.060 (see their Table 4, regression A).

presents the school participation time pattern for the full sample (with point estimates from Table 4, regression 5). The small but growing gaps between orphans and non-orphans during the two years prior to parent death are consistent with an extended period of parental morbidity. In regressions 3 and 5, F-tests reject the equality of parent death impacts three years before and three years after the death at 90 and 95 percent confidence, respectively. Effects are robust to the use of an alternative measure of school participation, the school enrollment indicator variable, yielding estimated magnitudes similar to Gertler et al. (2004) in Indonesia – more than a doubling of the drop-out rate after parent death – with time patterns similar to those for school participation (Appendix Table A1).

In terms of magnitudes, note that the estimated negative effect of a parent death on school participation (0.026-0.037), while not trivial, is far smaller in absolute value than the effect of eliminating a child's moderate-heavy worm infection (approximately 0.20, see Miguel and Kremer 2004). For another comparison, the impact of parent death is similar to the estimated impact of several poverty proxies: households without latrines (or toilets) at their home compound have school participation rates that are 0.031 lower than households with latrines, conditional on other household controls (Table 4, regression 1). In this sense, the findings are broadly consistent with much of the existing literature: the impact of parent death on school participation is indeed negative, but the average effect is moderate, even in an area of high HIV/AIDS prevalence – a finding that echoes work by Ainsworth et al. (2002). Note, however, that our parent death estimates may be lower bounds on true effects, as discussed above.

4.2 Impacts by Child, Parent, and Household Characteristics

The impacts of maternal and paternal death are nearly identical (Table 5, regression 1). The additional impact of becoming a double orphan, on top of the effect of losing both mother and father, on school participation is near zero and not statistically significant (regression 2), although there is limited statistical precision on this coefficient estimate as a result of the relatively small number of double orphans in the sample. Losing both parents has a negative impact somewhat larger than the effect of losing the first parent (regression 3), although we cannot reject that the effects of losing one versus both parents are equal at traditional confidence levels.[14] Among those who lost both parents, there is no significant difference between losing one's father first versus one's mother (regressions not shown).

Older orphans are often considered to be at higher risk of dropping out of school after parent death since the opportunity cost of their time is higher than for younger children, and some have claimed that girl children are also at higher risk of dropping out than boys (World Bank 2002, UNAIDS 2002). However, there are no significantly different impacts by child age or gender (Table 5, regressions 4 and 5), and these effects are reasonably precisely estimated.[15] Alternatively, it is possible that age is not as important as the relative age of the child in the household, i.e., older children may bear a greater burden following a parent death. Yet although there is no survey information on birth order in our dataset, the interaction of parent death and the number of infants in the child's household at baseline is not statistically significant (regressions not shown). Finally, the likelihood that an orphan is removed from school could in part depend on the child's expected returns to schooling, which is likely to be a function of ability; however, there is no statistically significant differential effect of parent death as a function of the child's class rank at baseline in 1998 (regressions not shown).

Parent death impacts differ across socioeconomic groups, with children from households with fewer assets experiencing the greatest reductions in school participation: children from households without latrines (or toilets) at home experience a drop in school participation of 0.090 (significant at 95 percent confidence, Table 6, regression 1) after parent death, while the effect for children from households with latrines is just -0.011 (not statistically significant). The pattern for coefficient estimates on other poverty proxies – cattle ownership, child nutrition, and an overall poverty index constructed

[14] The categories "Post-first parent death" and "Post-second parent death" in Table 5 are mutually exclusive: after a second parent death, the indicator variable "Post-first parent death" takes on a value of zero.

[15] Table 5, regression 4 also includes an indicator variable for "Missing age data", and interactions between the "Missing age data" indicator and the parent death indicators (coefficients not reported).

from these poverty proxies from the baseline surveys[16] – all yield similar, though not statistically significant effects, which taken together provides further evidence that children in initially poorer households are likely to be worse hit by parent death (regressions 2-4), while effects for better-off households are near zero. These findings suggest that moderate amounts of household wealth partially buffer children from the shock of parent death, or perhaps that households with greater wealth can call upon relatives with more resources to foster orphans, or to assist in the payment of school expenses, after the parent death.[17]

4.3 Impacts by Community Characteristics

Community characteristics are not associated with differential parent death impacts: most importantly, there is no statistically significant evidence that orphans fare worse in primary school communities with higher current rates of orphanhood (Table 7, regression 1), although the point estimate is reasonably large and in the expected direction, with larger school participation drops after parent death in communities with higher orphan rates. This main result on the local orphan rate is robust to an alternative definition of overall local orphan burden (regression 2), and to examining local maternal and paternal orphanhood separately (regression 3). Results are similar using initial 1998 orphan rates (regressions not shown).

The overall level of school participation is also not significantly lower in communities with higher rates of orphanhood, suggesting that non-orphans do not have substantially lower school participation in such communities, either. The large variation in orphan rates across communities in our study area makes it well-suited to explore this issue – there is considerable variation in baseline local

[16] We use principal components to construct an index of household assets (following Filmer and Pritchett 2001).

[17] Similarly the likelihood that a child is enrolled in school after a parent death falls 0.062 for children in households without latrines but only 0.015 for households with latrines, and falls 0.056 for children from households without cattle (about half of the sample), but near zero for households with cattle (regressions not shown). Patterns are similar for interactions with nutrition and the poverty index, and these coefficient estimates are somewhat more statistically significant when school enrollment is the dependent variable (regressions not shown).

orphan rates within our study area (Figure 2), from near zero in northern Funyula division to 40 percent in southern Budalangi.[18] Unfortunately, limited statistical precision means we cannot rule out moderate negative effects of the local orphan rate on the school participation of orphans.

These findings suggest that the claim that social networks in rural western Kenya are rapidly breaking down under the strain of HIV/AIDS deaths – and that as a result the growing numbers of orphans cannot be taken care of by surviving relatives and other community members – should be re-evaluated. Of course, further research is needed to understand how general these findings are beyond rural western Kenya, but the fact that there is little evidence networks are breaking down in this region, with its high orphan rates, may imply that this is even less of a concern in areas where rates are lower.

There are no significant differential effects of parent death on school participation related to the ethnic composition (Table 7, regression 4) of local orphans, or to the socioeconomic status of the community as a whole as proxied by average latrine ownership at baseline (regression 5).

4.4 Addressing Sample Attrition

We next place bounds on the impact of parent death on child school participation accounting for attrition and missing orphan status data. This exercise requires assumptions on three groups: (A) children who are missing orphan status data only, (B) children who are missing school participation data only (i.e., attrition), and (C) children who are missing both.

To establish an upper bound on impacts, we assume children in group A are orphans if their school participation is lower than the mean participation rate among children who do not become orphans during the period; otherwise we assume they are not orphans. For children in group B, we assume their school participation is equal to zero if they are orphans, and one if they are not orphans. Finally, we assume that all children in group C are orphans and that their school participation is equal to zero – an

[18] In contrast, Yamano and Jayne (2003) find that school attendance is strongly negatively correlated with lagged provincial HIV-prevalence rates in their sample.

extreme bounding procedure related to Manski (1995). (The observed timing of parent deaths across years for those whose orphan status is actually known is used to assign simulated years of parent death to the children assigned to be orphans in this bounding procedure.) Under these assumptions, and with 100 draws of the simulation in a specification with child fixed effects, the average upper bound is a 6.6 percentage point decrease in school participation following a parent death – a large but not massive effect.

The lower bound procedure makes polar opposite assumptions: group A children are assumed to be orphans if their school participation is higher than average, while for children in group B, school participation is assumed to be one if they are orphans, and zero otherwise. For group C, we assume that all are orphans and that all have perfect school participation. This yields a mean lower bound estimate of a 1.7 percentage point decrease in school participation – and thus even under these implausibly extreme conservative assumptions, we still estimate a small negative impact of parent death on schooling.

4.5 Macro-impacts of Parent Death on Future Income Levels

Most existing macroeconomic simulations predict moderate impacts of AIDS on economic growth: Bonnel (2000), Kambou et al. (1992), Over (1992), and Sackey and Raparla (2000) all find negative impacts on annual growth rates of less than 1.5 percent. However, these studies do not consider the human capital impacts of HIV/AIDS that are the focus of the current paper; Bell et al. (2003) do attempt to take into account the loss of human capital as a result of AIDS and predict major negative impacts for economies with "sufficiently high" disease prevalence.

This sub-section attempts to translate our microeconometric findings on parent death impacts into speculative estimated effects on lifetime income. First, orphans' decreased school participation is summed up over their entire childhood, under the following assumptions: first, the impact of parent death begins three years before the death, using the estimates reported in Table 4, regression 5; second, any lost schooling due to parent death only occurs before age 18, when individuals are assumed to leave school; and finally, the impact on school participation beyond three years after the parent death is equal to the effect in the third year after the death. Under these assumptions, a child who suffers a parent death at age

18 only experiences 0.16 fewer years of school participation (the sum of effects one, two, and three years before the death, and in the year of the death), while a child who experiences a parent death at age 5 loses 1.29 years using a similar calculation, including impacts in the years after the death. Median child age at parent death in the sample is 14 years, and these children lose 0.47 years of schooling on average due to parent death.

Knight and Sabot's (1990)[19] estimate for Kenya is used to translate the loss of schooling into earnings; they find the return to an additional year of primary school in Kenya is 7 percent. Output per worker in Kenya is US$570 (World Bank 1999). Following Miguel and Kremer (2004), we assume that 60 percent of output per worker is wages (yielding an annual wage of US$342); wage gains from increased school participation are earned over 40 years in the workforce; future wages are discounted at 5 percent per year; and there is no wage growth over time. Within this framework, the discounted value of future lost wages is US$194 for children orphaned at age 14, the median age at parent death, though note that the future income lost due to parent death is more than twice as high for children from poor households (e.g., without latrines) as for children from better off households.

The Knight and Sabot (1990) estimate allows us to compute a macroeconomic impact of parent death – typically due to HIV/AIDS in our study area – on income for the next generation, working through the schooling channel. Note that the proportion of children under age 15 who are orphans is thought to be approximately 12 percent in Kenya (UNAIDS 2002). Focusing on the median orphan as a benchmark, the estimated overall loss of future national income levels in Kenya as the result of parent

[19] Knight and Sabot (1990) estimate returns to education in Kenya controlling for a wide range of variables, including cognitive tests. They decompose returns to education into a return to cognitive performance (on tests of literacy, numeracy and reasoning) and a direct return to years of schooling and find that years of schooling alone accounts for approximately 0.4 of the total return to education. Knight and Sabot (1990) performed this decomposition for secondary education, but it serves as a useful approximation in the absence of a similar exercise for primary education. If one interprets the return to schooling as a human capital effect rather than a signaling effect, the return to an additional year of primary schooling is approximately (17 percent)*(0.4) = 7 percent.

death, given current orphan figures, is roughly (-0.47 years of school participation) x (7 percent return to a year of schooling) x (12 percent of the population are orphans) = 0.4 percent of national income. We also carry out a related calculation, taking into account the expected income effects for children who lose parents at different ages using the distribution of child age in our sample, as well as that primary school enrollment in Kenya is approximately 84 percent (African Development Bank 2001), and find an estimated loss of future national income of 0.5 percent (calculations not shown). The schooling channel, then, appears likely to have a minor impact on total future national income.

Still, it is worth noting that the negative impacts on individual income for particular sub-populations – in particular, poor orphans – are likely to be much larger, and the impact of the epidemic on future income would also rise if the proportion of orphans in the population continues to increase. Moreover, the estimates we present in this study miss effects of parent death on the schooling of children below age 5, as these children may never enroll in primary school in the first place. Finally, as with any microeconomic empirical estimates, the issue of generalizeability remains important, since the impact of parent AIDS deaths could differ across African settings – rural versus urban areas for instance, an issue we are unable to address in this study's entirely rural sample – and this issue urgently demands further micro-empirical investigation.[20]

5. Discussion

To summarize the main finding, parent death is associated with a drop in primary school participation in rural western Kenya, with impacts particularly large among households with fewer assets. The empirical approach addresses a number of shortcomings of existing studies, highlighting the usefulness of collecting panel data sets on the health and education of children in less developed countries.

---

[20] In future work, we plan to extend the analysis in this paper for several additional years using the Kenya Life Panel Survey, which will collect health, education, labor market, and demographic outcomes, including income, on the same sample of children, in order to estimate longer-term impacts of parent death on life outcomes.

The results also provide some additional insight into the debate over how to target assistance programs in areas with large numbers of AIDS orphans – and in particular into the issue of whether orphans should be specifically targeted for transfers, or whether such transfers should instead be directed to all individuals in communities with many orphans, or to poor children independent of orphan status. The results suggest that orphans are a particularly disadvantaged group in terms of school participation in rural Kenya, and that transfers should ideally be targeted to them directly. The finding that low socioeconomic status orphans are particularly disadvantaged means that easy markers of poverty could also be used to target assistance, in tandem with child orphan status, wherever possible. The finding that communities where high proportions of children are orphans do not have significantly lower average school participation suggests that targeting high prevalence areas might be less effective than targeting particular orphans in this context – although there may, of course, be program cost savings to targeting whole communities rather than individuals.

Finally, given the rapidly growing numbers of orphans in Sub-Saharan Africa, this paper is by no means the final word. Further research on the impact of parent HIV/AIDS deaths on children, as well as on the design of programs to assist orphans, in other African settings is desperately needed.

References

Adetunji, Jacob. (2000). “Trends in under-5 mortality rates and the HIV/AIDS epidemic”, *Bulletin of the World Health Organization*, 78, 1200-1206.

African Development Bank. (2001). *African Development Report 2001*. Oxford University Press.

Aigner, Dennis J., “Regression with a Binary Independent Variable Subject to Errors of Observation,” *Journal of Econometrics*, 1973, 1(1), pp. 49-50.

Ainsworth, Martha, Kathleen Beegle and Godlike Koda. “The impact of adult mortality on primary school enrolment in Northwestern Tanzania.” Africa Region Human Development Working Paper, World Bank, 2002.

Bell, Clive, Shantayanan Devarajan, and Hans Gersbach. “The long-run economic costs of AIDS: theory and an application to South Africa.” Unpublished manuscript, World Bank, 2003.

Bonnel, Rene. “HIV/AIDS: does it increase or decrease growth in Africa?” ACTAfrica, World Bank, November 2, 2000.

Case, Anne, Christina Paxson and Joseph Ableidinger. “Orphans in Africa: Parental Death, Poverty, and School Enrollment,” forthcoming *Demography*, 2004.

Deininger, Klaus, Marito Garcia, and K. Subbarao. “AIDS-induced orphanhood as a systemic shock: magnitude, impact, and program interventions in Africa.” *World Development*, July 2003, 31(7), 1201-20.

Filmer, Deon, and Lant Pritchett. “Estimating wealth effects without expenditure data – or tears: an application to education to educational enrollment in states of India.” *Demography,* 2001, 38, 115-32.

Foster, G., R. Shakespeare, F. Chinemana, H. Jackson, S. Gregson, C. Marange, S. Mashumba. “Orphans Prevalence and Extended Family Care in a Peri-urban Community in Zimbabwe”, *AIDS Care*, February 1995, 7(1), 3-18.

Foster, Geoff, and John Williamson. “A review of current literature on the impact of HIV/AIDS on children in Sub-Saharan Africa.” *AIDS*, 2000, 14 (suppl 3), S275-S284.

Gertler, Paul, David Levine and Minnie Ames. "Schooling and Parental Death", *Review of Economics and Statistics*, 86(1), February 2004.

Gertler, Paul, David Levine and Sebastian Martinez. "The presence and presents of parents: do parents matter for more than their money?" Unpublished manuscript, U.C. Berkeley, 2003.

Kamali, A., J. A. Seeley, A. J. Nunn, J. F. Kengeya-Kayondo, A. Ruberantwari, and D. W. Mulder. "The orphans problem: experience of a Sub-Saharan Africa rural population in the AIDS epidemic." *AIDS Care*, 1996, 8(5), 509-515.

Kambou, G., Shantayanan Devarajan, and Mead Over. "The economic impact of AIDS in an African country: simulations with a CGE model of Cameroon." *Journal of African Economies*, 1, 109-30.

Knight, J.B., and R.H. Sabot. *Education, Productivity, and Inequality*. Oxford University Press, Washington, D.C., 1990.

Lloyd, Cynthia, and Ann Blanc. "Children's schooling in Sub-Saharan Africa: the role of fathers, mothers, and others." *Population and Development Review, 1996,* 22(2), 265-298.

Manski, Charles. *Identification Problems in the Social Sciences*, Cambridge: Harvard Univ. Press, 1995.

Miguel, Edward, and Michael Kremer. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica*, 72(1), 159-217, 2004.

Miguel, Edward, and Mary Kay Gugerty. "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya", unpublished manuscript, University of California, Berkeley and University of Washington, 2002.

Morgan, Dilys, S. S. Malamba, J. Orem, B. Mayanja, M. Okongo, and J.A.G. Whitworth. "Survival by AIDS defining condition in rural Uganda," *Sexually Transmitted Infections*, 2000, 76, 193-97.

Morgan, Dilys, and Jimmy A.G. Whitworth. "The natural history of HIV-1 infection in Africa," *Nature Medicine*, February 2001, 7(2), 143-145.

NASCOP (National AIDS and STD's Control Programme). *Estimating National HIV Prevalence in Kenya from Sentinel Surveillance Data*, Nairobi, Kenya, June 2001.

Ntozi, James P.M. "Effect of AIDS on Children: The Problem of Orphans in Uganda", *Health Transition*

Review, 1997, supplement to volume 7, 23-40.

Over, A.M. “The macroeconomic impact of AIDS in Sub-Saharan Africa.” AFTPN Technical Working Paper 3, Washington, D.C., World Bank, 1992.

Robinson, Simon. “Orphans of AIDS,” *TIME*, December 13, 1999.

Ryder, Robert, Munkolenkole Kamenga, Muniaka Nkusu, Veronique Batter, and William Heyward. “AIDS orphans in Kinshasa, Zaire: incidence and socioeconomic consequences,” *AIDS*, December 13, 1999, 8, 673-9.

Sackey, James, and Tejaswi Raparla. “Lesotho: the development impact of HIV/AIDS: selected issues and options.” World Bank Report No. 21103-LSO, 2000.

UNAIDS, UNICEF, and USAID. Children on the Brink 2002: a joint report on orphan estimates and program strategies. 2002.

UNAIDS. *Report on the Global HIV/AIDS Epidemic*. (www.unaids.org), 2002.

United Nations Development Programme (UNDP). *Human Development Indicators 2002*.

Watson, Carol. “RAPing in Chad.” *Rapid Assessment Procedures: Qualitative Methodologies for Planning and Evaluation of Health Related Programmes.* Eds. N.S. Scrimshaw and G.R. Gleason. International Nutrition Foundation for Developing Countries. Boston: 409-16, 1992.

Wax, Emily. “A Generation Orphaned by AIDS,” *Washington Post*, August 13, 2003, pp. A01.

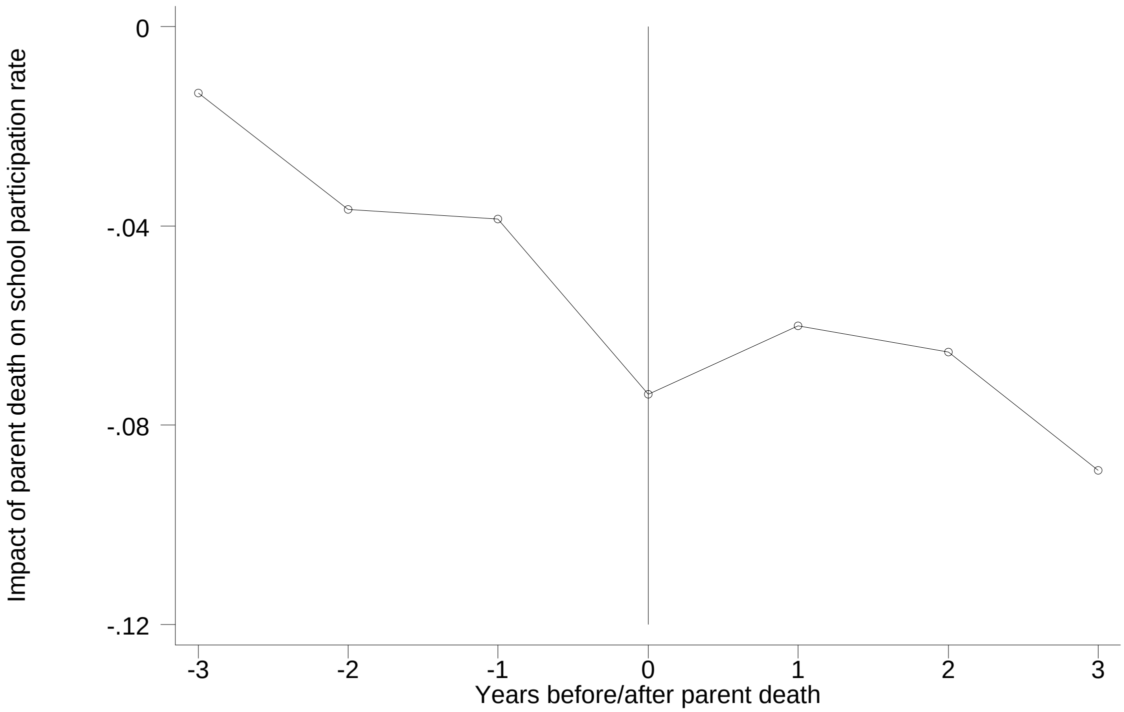
World Bank. World Development Indicators (*www.worldbank.org*), 1999.

World Bank. *Education and HIV/AIDS: A Window of Hope*. Washington DC: The World Bank, 2002.

Yamano, Takashi, and T.S. Jayne. “Measuring the impact of working-age adult mortality on small-scale farm households in Kenya,” *World Development*, January 2004, 32(1), 91-119.

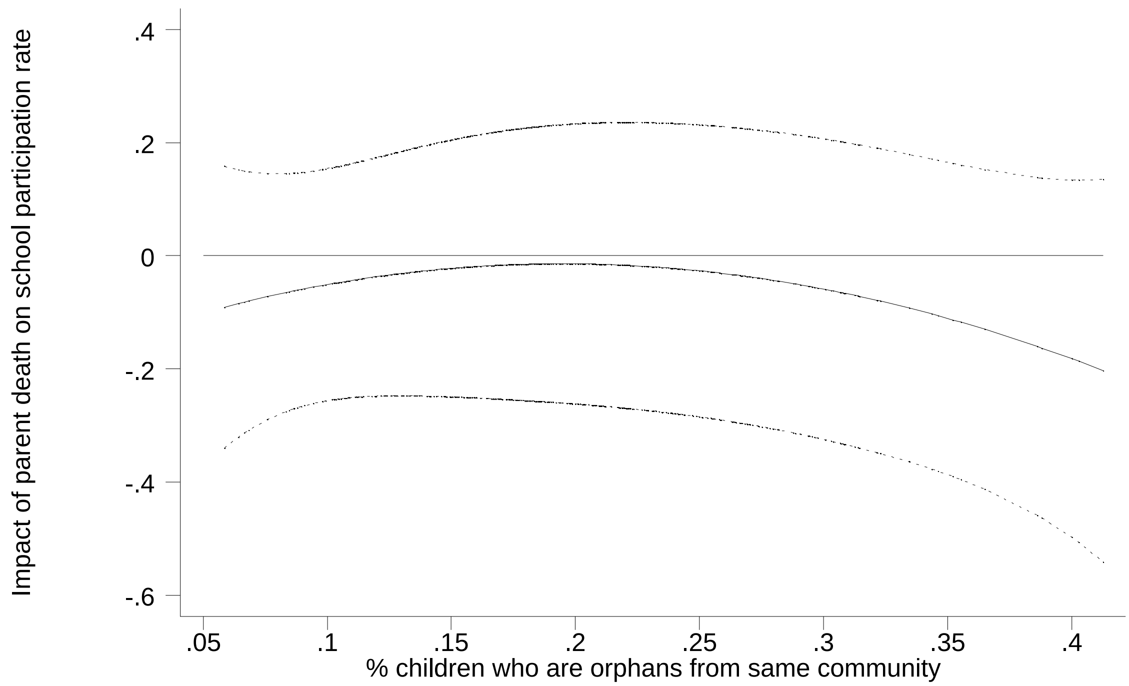
Yamano, Takashi, and T.S. Jayne. “Working-age adult mortality and primary school attendance in rural Kenya,” *FASID Discussion Paper Series on International Development Strategies*, No. 2003-006, 2003.

<u>Figure 1:</u> Parent death and school participation over time (relative to four years prior to parent death)



<u>Notes:</u> Full sample, point estimates from Table 4, regression 5.

<u>Figure 2:</u> The estimated impact of parent death on orphan school participation, as a function of proportion of orphans in the school community



<u>Notes:</u> Cubic regression fit, dashed lines are 95% confidence intervals. Results are for the full sample. The sample of children for determining the variable on the horizontal axis is all children from the initial sample for whom orphan status is available, whether still enrolled in school or not.

Table 1: Summary statistics

	Obs	Mean	Std. Dev	Min	Max
Panel A: Full sample					
Became orphan, during 1999-2002	18133	0.08	0.27	0	1
School participation, 1998	18133	0.85	0.23	0	1
School enrollment, 1998	18133	0.98	0.14	0	1
Female	18133	0.48	0.5	0	1
Age, 1998	14970	11.79	2.53	5	18
Proportion orphans in school, 1998	18133	0.15	0.05	0.01	0.41
Proportion maternal orphans in school, 1998	18133	0.05	0.02	0.01	0.20
Proportion paternal orphans in school, 1998	18133	0.12	0.04	0	0.33
Proportion double orphans in school, 1998	18133	0.02	0.01	0	0.12
Panel B: Restricted sample					
Became orphan, 1999-2002	7815	0.09	0.28	0	1
School participation, 1998	7815	0.92	0.17	0	1
School enrollment, 1998	7815	1	0.06	0	1
Female	7815	0.48	0.50	0	1
Age, 1998	7769	12.91	2.00	6	18
Proportion orphans in school, 1998	7815	0.14	0.05	0.01	0.35
Proportion maternal orphans in school, 1998	7815	0.05	0.02	0.01	0.15
Proportion paternal orphans in school, 1998	7815	0.11	0.04	0	0.31
Proportion double orphans in school, 1998	7815	0.02	0.01	0	0.12
Child weight-for-age (z-score), 1998	7815	-1.44	0.82	-4.79	2.34
Child had malaria/fever in past month, 1998	7815	0.39	0.49	0	1
Child wears shoes, 1998	7815	0.14	0.35	0	1
Child wears school uniform, 1998	7815	0.86	0.34	0	1
Child appears "clean", 1998	7815	0.62	0.49	0	1
Latrine at home, 1998	7815	0.82	0.38	0	1
Cows at home, 1998	7815	0.49	0.50	0	1
Goats at home, 1998	7815	0.41	0.49	0	1
Poultry at home, 1998	7815	0.93	0.25	0	1

Notes: School participation variables are from regular unannounced attendance checks collected throughout the 1998 to 2002 school years (see Miguel and Kremer 2004). Orphan status variables are from the 2002 Tracking Data. Demographic and socioeconomic characteristics are from the 1998 Pupil Questionnaire. The reduced samples for "age" is due to missing data; in the regressions, we include an indicator variable for missing age data.

<u>Table 2:</u> Attrition and Child Characteristics

	Dependent variable:			
	Missing parent death information		Attrited during 1998-2002 (missing school participation)	
	Probit	Probit	Probit	Probit
	(1)	(2)	(3)	(4)
Became orphan, 1999-2002			-0.056***	-0.053***
			(0.011)	(0.016)
Female	0.029**	0.034	-0.006	-0.281***
	(0.012)	(0.063)	(0.016)	(0.057)
Age in 1998	0.016***	0.020***	0.007***	-0.012
	(0.002)	(0.007)	(0.001)	(0.008)
Missing age data indicator	0.399***		0.083***	
	(0.033)		(0.025)	
Female * Age in 1998	0.000	-0.001	0.001*	0.021***
	(0.001)	(0.005)	(0.002)	(0.005)
School in Budalangi Division	0.073***	0.065***	0.008	0.004
	(0.018)	(0.018)	(0.013)	(0.016)
Child weight-for-age (z-score), 1998		-0.074***		0.001
		(0.025)		(0.035)
Child had malaria/fever in past month, 1998		0.014*		-0.010
		(0.009)		(0.009)
Latrine at home, 1998		-0.063		0.126*
		(0.054)		(0.060)
Cows at home, 1998		-0.107*		-0.004
		(0.057)		(0.071)
Goats at home, 1998		-0.018**		-0.022*
		(0.008)		(0.012)
Poultry at home, 1998		-0.037**		-0.004
		(0.016)		(0.022)
Child wears shoes, 1998		0.033***		0.051**
		(0.012)		(0.022)
Child wears school uniform, 1998		-0.026*		0.050***
		(0.014)		(0.014)
Child appears 'clean,' 1998		-0.030		0.340***
		(0.054)		(0.048)
Cows at home, 1998 * Age in 1998		0.007		-0.001
		(0.004)		(0.005)
Latrine at home, 1998 * Age in 1998		0.005		-0.008
		(0.004)		(0.006)
Child appears 'clean', 1998 * Age in 1998		0.002		-0.025***
		(0.004)		(0.005)
Child weight-for-age (z-score), 1998 * Age in 1998		0.007		-0.001
		(0.002)		(0.003)
Observations	24,111	9,789	18,133	7,815
Mean (s.d.) of dependent variable	0.25	0.20	0.23	0.28
	(0.43)	(0.40)	(0.42)	(0.45)

<u>Notes:</u> All regressions are probits, with marginal effects reported. Standard errors are clustered at the school level. * significant at 90%; ** 95%; *** 99% confidence. Unreported controls include deworming program (PSDP) treatment group indicator variables, region-year indicator variables, and the constant. Regression 1 includes all 24,111 children from the baseline sample, and regression 2 includes all baseline children for whom covariates are available. Regression 3 includes all 18,133 children for whom parent mortality data is available, and regression 4 includes those among the 18,133 for whom covariates are available. Age data is missing for 5,095 baseline children.

Table 3: Baseline characteristics for children who lost a parent versus those who did not

	Became orphans	Never orphans	Difference Became – Never (s.e.)	Orphan status unknown
Panel A: Full sample				
School participation, 1998	0.87	0.87	0.00 (0.01)	0.76
School enrollment, 1998	0.99	0.99	0.00 (0.00)	0.93
Female	0.46	0.48	-0.02 (0.02)	0.51
Age	11.78	11.79	-0.01 (0.07)	12.35
N	1245	13725		4046
Panel B: Restricted sample				
School participation, 1998	0.92	0.92	0.00 (0.01)	0.87
School enrollment, 1998	1.00	1.00	0.00 (0.00)	0.99
Female	0.48	0.48	0.00 (0.02)	0.53
Age	12.74	12.92	-0.19** (0.08)	13.53
Child weight-for-age (z-score), 1998	-1.40	-1.45	-0.04 (0.03)	-1.34
Child had malaria/fever in past month, 1998	0.40	0.39	0.01 (0.02)	0.42
Child wears shoes, 1998	0.13	0.14	-0.01 (0.01)	0.19
Child wears school uniform, 1998	0.85	0.86	-0.02 (0.01)	0.85
Child appears "clean", 1998	0.59	0.62	-0.03* (0.02)	0.64
Latrine at home, 1998	0.81	0.82	-0.01 (0.02)	0.81
Cows at home, 1998	0.49	0.49	0.00 (0.03)	0.44
Goats at home, 1998	0.39	0.41	-0.02 (0.02)	0.37
Poultry at home, 1998	0.93	0.93	0.00 (0.01)	0.91
N	667	7148		1938
Panel C: Subsample of children with school participation data in 1997 and 1998				
School participation, 1997	0.84	0.81	0.03 (0.02)	0.75
School participation, 1998	0.80	0.79	0.01 (0.02)	0.67
N	250	2673		904

Notes: Standard errors are clustered at the school level. * significant at 90%; ** 95%; *** 99% confidence. For Panel C, 27 schools are included which were involved in another NGO program, and thus had 1997 attendance data. The reduction in sample size in Panel A is due to missing age information in the Full sample. The final column includes children who would be in the full sample (or restricted sample) but for the lack of parent mortality data.

Table 4: Impact of parent death on school participation

	Dependent variable: Total school participation				
	OLS Restricted sample (1)	OLS Restricted sample (2)	OLS Restricted sample (3)	OLS Full sample (4)	OLS Full sample (5)
Post-parent death	-0.037***	-0.026*		-0.036***	
	(0.012)	(0.014)		(0.009)	
3 years pre-death			0.007		-0.013
			(0.036)		(0.026)
2 years pre-death			-0.037		-0.037
			(0.038)		(0.029)
1 year pre-death			-0.037		-0.039
			(0.041)		(0.030)
Year of parent death			-0.054		-0.074**
			(0.041)		(0.031)
1 year post-death			-0.054		-0.060**
			(0.043)		(0.030)
2 years post-death			-0.071		-0.065**
			(0.049)		(0.033)
3 years post-death			-0.070		-0.089**
			(0.059)		(0.040)
Child weight-for-age (z-score), 1998	-0.009**				
	(0.003)				
Child had malaria/fever in past month, 1998	-0.017***				
	(0.005)				
Child wears shoes, 1998	0.019*				
	(0.011)				
Child wears school uniform, 1998	0.047***				
	(0.007)				
Child appears 'clean,' 1998	0.021***				
	(0.005)				
Latrine at home, 1998	0.031***				
	(0.006)				
Cows at home, 1998	0.022***				
	(0.005)				
Goats at home, 1998	0.003				
	(0.006)				
Poultry at home, 1998	0.004***				
	(0.009)				
Student fixed effects	No	Yes	Yes	Yes	Yes
Observations	30817	30817	30817	73070	73070
Mean (s.d.) of dependent variable	0.77	0.77	0.77	0.75	0.75
	(0.34)	(0.34)	(0.34)	(0.33)	(0.33)
R^2	0.11	0.58	0.58	0.54	0.54

Notes: All regressions are ordinary least squares. Standard errors are clustered at the school level. * significant at 90%; ** 95%; *** 99% confidence. Unreported controls include deworming program (PSDP) treatment group indicator variables, birth-year cohort-year-gender indicator variables, and region-year indicator variables; and the constant term. Regressions 1, 2 and 3 contain 7,815 unique pupils. Regressions 4 and 5 contain unique 18,133 pupils.

Table 5: Impact of maternal and paternal death by child gender and age

	Dependent variable: Total school participation				
	OLS Full sample (1)	OLS Full sample (2)	OLS Full sample (3)	OLS Full sample (4)	OLS Full sample (5)
Post-maternal death	-0.036**	-0.039**		-0.033*	-0.028
	(0.015)	(0.016)		(0.020)	(0.020)
Post-paternal death	-0.032***	-0.034***		-0.028*	-0.025*
	(0.011)	(0.012)		(0.016)	(0.014)
Post-maternal death * Post-paternal death		0.015			
		(0.038)			
Post-first parent death			-0.035***		
			(0.010)		
Post-second parent death			-0.058*		
			(0.033)		
Child below age 12 * Post-maternal death				-0.006	
				(0.025)	
Child below age 12 * Post-paternal death				0.009	
				(0.020)	
Female * Post-maternal death					-0.018
					(0.032)
Female * Post-paternal death					-0.015
					(0.020)
Student fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	73070	73070	73070	73070	73070
Mean (s.d.) of dependent variable	0.75	0.75	0.75	0.75	0.75
	(0.33)	(0.33)	(0.33)	(0.33)	(0.33)
R^2	0.54	0.54	0.54	0.54	0.54

Notes: All regressions are ordinary least squares. Standard errors are clustered at the school level. * significant at 90%; ** 95%; *** 99% confidence. Unreported controls include deworming program (PSDP) treatment group indicator variables, birth-year cohort-year-gender indicator variables, region-year indicator variables, and the constant term. The categories “Post-first parent death” and “Post-second parent death” are mutually exclusive: after a second parent death, the indicator variable “Post-first parent death” takes on a value of zero. Regression 4 also includes the Missing age data indicator variable, and interactions between the Missing age data indicator and (Post-maternal death) and with (Post-paternal death) – coefficient estimates not reported. All regressions contain 18,133 unique pupils.

Table 6: Impact of parent death by child baseline socioeconomic status

	Dependent variable: Total school participation			
	OLS Restricted sample (1)	OLS Restricted sample (2)	OLS Restricted sample (3)	OLS Restricted sample (4)
Post-parent death	-0.011	-0.020	-0.021	-0.018
	(0.015)	(0.017)	(0.017)	(0.016)
Post-parent death * No latrine at home	-0.079**			
	(0.036)			
Post-parent death * No cows at home		-0.012		
		(0.024)		
Post-parent death * Undernourished child			-0.023	
			(0.032)	
Post-parent death * Poor household				-0.040
				(0.038)
Student fixed effects	Yes	Yes	Yes	Yes
Observations	30817	30817	30817	30817
Mean (s.d.) of dependent variable	0.77	0.77	0.77	0.77
	(0.34)	(0.34)	(0.34)	(0.34)
R^2	0.58	0.58	0.58	0.58

Notes: All regressions are ordinary least squares. Standard errors are clustered at the school level. * significant at 90%; ** 95%; *** 99%. Unreported controls include deworming program (PSDP) treatment group indicator variables, birth-year cohort-year-gender indicator variables, region-year indicator variables, and the constant term. "Undernourished" is an indicator variable that takes on a value of one if the weight-for-age z-score is less than -2. "Poor" is an indicator variable that takes on a value of one for students whose households are in the bottom quintile of a poverty index; the index is created through a principal components methodology, and the inputs are the household socioeconomic measures (latrine ownership, cow ownership, goat ownership, poultry ownership, child wears shoes, child wears school uniform, child is "clean"). All regressions contain 7,815 unique pupils.

Table 7: Community characteristics and the impact of parent death

	Dependent variable: Total school participation				
	OLS Full sample (1)	OLS Full sample (2)	OLS Full sample (3)	OLS Restricted sample (4)	OLS Full sample (5)
Post-parent death	0.030	-0.036*	0.036	-0.006	-0.017
	(0.048)	(0.019)	(0.044)	(0.045)	(0.027)
Proportion orphans in school	-0.305				
	(0.335)				
Post-parent death * Proportion orphans in school	-0.272 (0.211)				
School above 75th percentile in proportion orphans		0.006			
		(0.017)			
Post-parent death * School above 75th percentile in proportion orphans		0.000 (0.020)			
Proportion maternal orphans in school			0.090		
			(0.486)		
Proportion paternal orphans in school			-0.313		
			(0.331)		
Post-parent death * Proportion maternal orphans in school			-0.002 (0.347)		
Post-parent death * Proportion paternal orphans in school			-0.378 (0.268)		
Proportion orphans in own language group				-0.032	
				(0.136)	
Proportion orphans in other language groups				-0.042	
				(0.060)	
Post-parent death * Proportion orphans in own language group				-0.001 (0.159)	
Post-parent death * Proportion orphans in other language groups				-0.057 (0.072)	
Post-parent death * Proportion households in community with no latrine					-0.087 (0.111)
Student fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	73070	73070	73070	27267	73070
Mean (s.d.) of dependent variable	0.75	0.75	0.75	0.77	0.75
	(0.33)	(0.33)	(0.33)	(0.34)	(0.33)
R^2	0.58	0.58	0.58	0.63	0.58

Notes: All regressions are ordinary least squares. Standard errors are clustered at the school level. * significant at 90%; ** 95%; *** 99%. Unreported controls include deworming program (PSDP) treatment group indicator variables, birth-year cohort-year-gender indicator variables, region-year indicator variables, and the constant term. Regressions 1, 2, 3 and 5 contain 18,133 unique pupils, and regression 4 contains 6,929 unique pupils, with the reduced sample size a result of missing child ethno-linguistic group data.

Appendix

Table A1: Impact of parent death on school enrollment

	Dependent variable: Total school enrollment				
	OLS Restricted sample (1)	OLS Restricted sample (2)	OLS Restricted sample (3)	OLS Full sample (4)	OLS Full sample (5)
Post-parent death	-0.026**	-0.024**		-0.026***	
	(0.010)	(0.012)		(0.008)	
3 years pre-death			0.002		-0.025
			(0.026)		(0.020)
2 years pre-death			-0.012		-0.030
			(0.029)		(0.021)
1 year pre-death			-0.012		-0.036
			(0.032)		(0.023)
Year of parent death			-0.027		-0.055**
			(0.032)		(0.024)
1 year post-death			-0.041		-0.056***
			(0.034)		(0.022)
2 years post-death			-0.055		-0.064***
			(0.038)		(0.023)
3 years post-death			-0.065		-0.090**
			(0.053)		(0.033)
Child weight-for-age (z-score), 1998	-0.015***				
	(0.003)				
Child had malaria in past month, 1998	-0.006				
	(0.004)				
Child wears shoes, 1998	-0.005				
	(0.008)				
Child wears school uniform, 1998	0.019***				
	(0.006)				
Child appears ‘clean,’ 1998	0.005				
	(0.003)				
Latrine at home, 1998	0.009*				
	(0.005)				
Cows at home, 1998	0.010**				
	(0.004)				
Goats at home, 1998	0.008**				
	(0.004)				
Poultry at home, 1998	-0.003				
	(0.007)				
Student fixed effects	No	Yes	Yes	Yes	Yes
Observations	30817	30817	30817	73070	73070
Mean (s.d.) of dependent variable	0.88	0.88	0.88	0.89	0.89
	(0.33)	(0.33)	(0.33)	(0.32)	(0.32)
R^2	0.11	0.53	0.53	0.51	0.51

Notes: All regressions are ordinary least squares (linear probability model). Standard errors are clustered at the school level. * significant at 90%; ** 95%; *** 99% confidence. Unreported controls include deworming program (PSDP) treatment group indicator variables, birth-year cohort-year-gender indicator variables, and region-year indicator variables; and the constant term. Regressions 1, 2 and 3 contain 7,815 unique pupils. Regressions 4 and 5 contain 18,133 unique pupils.