INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2010-53

Death and Schooling Decisions over the Short and Long Run in Rural Madagascar

J.-N. SENNE[1]

1 Centre de Recherche en Economie et Statistiques (CREST), Paris School of Economics (PSE) – UMR DIAL-Dauphine.
Address CREST : 15 Boulevard Gabriel Péri, 92245 Malakoff, France. Tel : +33 (01) 41 17 60 37. Email : jean-noel.senne@ensae.fr

Death and Schooling Decisions over the Short and Long Run in Rural Madagascar

Jean-Noël SENNE*

August 2010

Abstract

This paper provides strong evidence that adult mortality has a negative impact on children education outcomes, both over the short and the long run, in rural Madagascar. The underlying longitudinal data set and the difference-in-differences strategy used overcome most of the previous cross-section studies limitations, such as failure to control for child and household pre-death characteristics and unobserved heterogeneity. This paper also pays special attention to the heterogeneity and robustness of the effects estimated. Using a three year panel of school-aged children, our results show that orphans are 20% less likely to attend school the year following death than their non-orphaned counterparts. This effect is even more pronounced for girls, young orphans and children from relatively poorer households. Pushing further the analysis to a sample of adults, our results show that those who became orphans in their childhood completed on average one year of education less. These findings suggest that, in a context where resources are scarce and formal insurance and market mechanisms are failing, not only do households suffering unexpected shocks resort to children schooling adjustments as an immediate risk coping strategy, but also that adversity has long-lasting effect on human capital accumulation.

KEYWORDS : adult mortality, orphans, education, panel data

JEL - CODES : I10, I20, I31

*Centre de Recherche en Economie et Statistiques (CREST-INSEE) - Paris School of Economics (PSE) - UMR DIAL-Dauphine. Address: 15 Boulevard Gabriel Péri - 92245 Malakoff - (France), Phone: +33 1 41176037, E-mail: jean-noel.senne@ensae.fr

1 Introduction

The question of the impact of adult mortality on children schooling decisions is becoming a major concern, especially in sub-Saharan Africa. The spread of the AIDS epidemic in the recent decades seriously increased the adult death rate and consequently the number of young orphans. In sub-Saharan African, an estimated 10% of children under 15 have lost at least one parent (Hunter & Williamson, 2000). This has raised serious questions about the link between adult death and schooling investments for those children left behind. Indeed, if the loss of a parent can have a direct short-term impact on the subsequent schooling decisions, there could also be considerable long-term effects in terms of human capital accumulation in the long-run and earnings prospects for future generations. Shortfalls in human capital investments in children could severely endanger the growth potential of developing economies.

The international community is showing a growing concern in this issue, so that several programs were especially designed to foster orphans education[1]. However, these programs rely on weak empirical evidence, partly due to the lack of suitable data. Moreover, the few available empirical studies point out quite mixed conclusions. Yet properly addressing this issue is crucial for policy recommendations. Indeed, if orphanhood has no pure effect on schooling, there is no reason why orphans education should be especially targeted. Education programs could be designed on other observable criteria, such as poverty, directly linked to schooling outcomes and potentially correlated with death. On the contrary, if, everything being equal, orphans are at greater risk, there is an urging demand for education support in this population. That is why clearly understanding orphans vulnerability remains essential.

This paper, based on a representative sample of malagasy rural households, provides

[1]Many national and international programs were launched by international organizations such as the World Bank or UN agencies, in partnership with community-based associations and NGOs, to reduce school fees and expenses, to supply uniforms, improve access to credit or promote part-time education for orphans (see for instance the UNESCO program to Provide Education to Orphans and Vulnerable Children, the Orphan Support Africa program or the numerous National Orphan programs in several countries from East and Southern Africa).

strong evidence of a negative impact of adult death on subsequent children (aged 6-18) schooling decisions, both over the short and the long run. Unlike most previous studies which focus on parental deaths in highly HIV-infected areas, it investigates the impact of any adult death occurring in the household, regardless of its cause[2]. The underlying longitudinal data set allows us to overcome most of the cross-section studies limitations, such as failure to control for child and household pre-death characteristics and unobserved heterogeneity. Primarily using a three year panel of school-aged children and building our identification strategy on difference-in-differences methods, we show that orphans are 20% less likely to attend school the year following death than their non-orphaned counterparts. The effect is even more pronounced for girls, young orphans and children from relatively poorer households. This heterogeneity of impact according to gender, age and initial household characteristics gives insights about the underlying mechanisms through which death affects education outcomes. Pushing further the analysis to a sample of adults with completed schooling and known orphan history, we find that adults who experienced a parental death in their childhood completed on average one year of education less. These findings suggest that, in a context where resources are scarce and formal insurance and market mechanisms are failing, not only do households suffering unexpected shocks resort to children schooling adjustments as an immediate risk coping strategy, but also that adversity has long-lasting effect on human capital accumulation.

In what follows, Section 2 reviews the existing theoretical and empirical literature on adult mortality and schooling outcomes. Section 3 presents the data sets and descriptive statistics related to this issue in the malagasy context. Section 4 gives the econometric specification of the equations to be estimated and provides results. Section 5 assesses the robustness of estimations with respect to selection issues. Section 6 investigates the heterogeneity of effects and long-run persistence. Finally, Section 7 concludes.

[2]For the sake of simplicity, any child experiencing an adult death in the household will be referred to as an "orphan" throughout this paper, even if he (resp. she) is not strictly the son (resp. daughter) of the deceased member. This is not so restrictive in the malagasy context where most households are nuclear families. Most adult deaths registered in data are then affecting parents, and if not, a very close relative.

2 Death and schooling : a literature review

The existing literature related to the issue of the impact of adult mortality on education outcomes is mainly empirical. Most studies focus on the estimation of the overall sign and size of a potential effect. However, despite the difficulty to build a formal theoretical model, turning to the theoretical literature on human capital investments, intergenerational altruism and intrahousehold allocation is quite instructive to identify channels through which adult death could affect children schooling decisions.

2.1 Conceptual framework

The model of investments in children's human capital developed by Becker & Tomes (1979) in their seminal paper provides a useful framework to understand how schooling decisions could be altered in households experiencing death. Based on intergenerational altruism, the model derives the result that a family anticipating its future children's earnings realize the optimal investment which equalizes the marginal returns to education to the marginal costs. This well-known result relies on the three following assumptions, i.e (i) schooling is an investment exclusively valued through its contribution to future earnings, (ii) capital markets are perfect and households can always borrow against their children's anticipated incomes, and (iii) households' preferences are egalitarian among children and the opportunity cost of children's time is not affected by shocks such as death. In this world, households are neither liquidity nor time-constrained. Thus an important feature of the model is that investments in children's human capital are unresponsive to shocks to current financial and human households' resources. However, as noted by Gertler et al. (2004), these assumptions are unlikely to hold in the context of developing countries. Relaxing them exhibits several pathways by which death could affect schooling.

First, adult death credibly constitutes a large income shocks at the household level. On the one hand, medical expenditures associated with severe illness as well as funeral costs can be considerably high. This is particularly true in Madagascar, where funeral customs are numerous and a fundamental event in social life. On the other hand, death can induce an heavy income shortfall if it affects a working-age member. If resources become scarce and in a context of restricted access to credit and insurance, households who are liquidity constrained may be unable to bear subsequent education costs and may delay or discontinue children's schooling. Investments remain unaffected by a shock only if households have sufficient assets and/or precautionary savings to cope with.

Secondly, households experiencing adult illness and death face an additional demand for caregiving, domestic and productive labor. The implied increase in the opportunity cost of children's time may prompt households to withdraw them from school. Yet, shortage of intern labor previously supplied by the deceased member may be partially compensated by hiring workers or welcoming new members (Beegle, 2003; Yamano & Jayne, 2004). This is less likely to happen in rural areas where human and financial resources are scarce and where child labor is a natural substitute for domestic and productive adult labor.

Finally, even in the absence of financial and labor constraints, households' preferences and education promotion may be altered after the occurrence of an adult death. Indeed, the loss of a member reduces the total time spent in household activities, such as educational support. Besides, some authors contend that high mortality rates and decline in observed life expectancy affects households' expectations about lifetime returns to children's education (McPherson, 2001). Moreover, the loss of a close relative may be a traumatic event which affects achievement in school. On the whole, these combined factors suggest that death might diminish the expected returns to schooling and consequently human capital investments in children. If household's time and social norms are crucial inputs in the education production function, such effects may be non negligible.

Though the above mentioned pathways suggest an overall adverse effect of adult mortality on schooling, it is important to note that this impact might not be homogenous between and within households. Its magnitude hinges upon the vulnerability of households to shocks, which in turn depends on their initial resources and/or their ability to call on insurance mechanisms[3]. Besides, effects can be heterogenous along children's observable characteristics, such as age, gender or link to the deceased member.

Generally, we lack suitable instruments and data to specifically identify which pathway prevails. Therefore, most empirical studies focus on the estimation of the cumulative effect via all channels. Yet, if we believe in the income shock story, we expect the impact to be more acute in poorer households and in households experiencing a male death, men being the main income providers. If we believe in the opportunity cost story, we expect girls and older children to share a larger part of the burden as they are more likely to undertake domestic and productive tasks in those rural areas. Finally, if returns to education are taken into account, boys and older children should be relatively favored. That is why our estimations will pay special attention to the heterogeneity of impact according to age, gender, initial wealth and household composition, which gives insights into the underlying mechanisms at work.

2.2 Empirical evidence

The very first empirical studies that deal with the issue of adult mortality in the era of the HIV/AIDS epidemic addressed its impact on households' wealth and showed strong evidence that death corresponds to a large and permanent shock on income and consumption (Yamano & Jayne, 2004; Naidu & Harris, 2005; Beegle et al., 2006). Since then, the empirical literature has rather focused on the consequences of orphanhood on health and schooling outcomes. Though rich and prolific, this literature is mainly based on cross-sectional data, which so far yielded quite mixed results.

The seminal study on sub-Saharan Africa was conducted by Case et al. (2004) on large Demographic and Health Survey (DHS) nationally representative samples, covering

[3]Several studies point out the importance of informal insurance in developing countries where formal mechanisms are failing. This includes risk-sharing systems within extended families or community networks, likely to provide financial, human or moral assistance and thus working as safety nets (Townsend, 1995). Likewise, child fostering is a common practice among orphans. Akresh (2007) show that foster children are more likely to be enrolled in school years after death than their non-foster siblings. However, the same authors put forward the reciprocity involved in those mechanisms, which is hardly sustainable in case of large and permanent shocks such as death (Townsend, 1994).

10 countries. They show that, controlling for a whole set of household and child characteristics, primary school enrollment rates among orphans was significantly lower in most of the countries considered. However, a similar concurrent study by Ainsworth & Filmer (2002) on 28 countries, mostly in Africa, do not prove so conclusive and indicates that this difference varies greatly according to differential wealth across and within countries. It echoed several previous and following studies on more restricted areas (Ryder et al., 1994; Kamali et al., 1996; Lloyd & Blanc, 1996; Bicego et al., 2003; Ainsworth et al., 2005) and international organization reports (UNAIDS, 2002; Bank, 2002) which observe an impact of parents death only in poor households, particularly for female, young and maternal orphans whose enrollment is either discontinued or delayed.

One explanation for the mitigated impact found in those studies is that family and community networks work as efficient insurance systems, through child fostering and orphans caregiving (Foster et al., 1995; Akresh, 2007). Alternatively, death can be highly correlated with the socioeconomic status of the household, which is also a strong determinant of the demand for schooling. This puts forward one general limitation of cross-sectional studies which identify the effect of orphanhood only controlling for current household and child characteristics, that is to say after death occurred. However, current orphan socioeconomic status and living arrangements may themselves have been affected by death and may thus confound its true effect, since children now differs along observable characteristics potentially correlated with their orphan status. In other words, such studies fail to compare children with similar baseline characteristics, prior to an adult death. Another major concern is that, to the extent that factors conjointly influencing death and schooling are unobserved, such estimations may suffer from omitted variable bias.

Due to the lack of suitable data, fewer studies use longitudinal data to overcome endogeneity and omitted variable issues. Case & Ardington (2006) and Evans & Miguel (2007), respectively using a 2 year panels from South Africa and Kenya, implement household fixed-effect estimations to assess the impact of orphanhood on children's education outcomes, such as school enrollment and school participation. They find a significant negative effect only for maternal deaths. Evans & Miguel (2007) also show that this effect predates death, probably owing to the demand for caregiving of ill members, and that omitted variables tend to bias the estimates downward in cross-sectional studies. (Yamano & Jayne, 2005) find similar evidence from another 3 year panel of kenyan households, though analyzing the broader issue of working-age adult mortality. They point out the fact that the adverse effect identified is mostly driven by poorer households, girls before death and boys after death.

This paper builds on this recent empirical work to provide strong evidence that adult mortality has adverse effects on subsequent children schooling decisions. Following Yamano & Jayne (2005) and Evans & Miguel (2007) difference-in-differences framework, we evaluate in the same way the immediate impact of any adult death, considering that any adult household member might play a role in the education production function, regardless of its age and link to the surviving children. We use data from a 3 year (2004-2005-2006) panel survey on a representative sample of malagasy rural households. The survey design was made to register each death occurring in the household between the two rounds, allowing us to control for a wide set of characteristics prior to death, as well as unobserved heterogeneity. Yet, unlike previous longitudinal studies, we investigate more precisely the robustness and heterogeneity of effects across orphans. Above all, we also provide further unique evidence that orphanhood matters in the long-run. Utilizing a specific retrospective section on life history of the 2004 survey round, we find that adults who lost a parent as a child are worse off in terms of human capital accumulation.

3 Data and Preliminary Analysis

3.1 Survey Design

This paper uses data from an original longitudinal household survey implemented by the "Réseaux des Observatoires Ruraux malgaches" (ROR). This survey aims at illustrating specific issues regarding rural Madagascar with a focus on poverty dynamics in the long-run. The project was first initiated in 1995 with 4 rural observatories. The scope of the survey was gradually extended to new observatories all over the country to cover the diversity of agroecological areas and to collect data on a nationally representative sample of rural households. In the 2004, 2005 and 2006 rounds that we use, there is a total of 10 different observatories.[4] Within each of them, a minimum of 500 households are randomly selected among villages after an exhaustive population census. There are then traced each subsequent year, and surveyed if they were correctly re-contacted or replaced if they were not found, to keep the sample size relatively constant. The baseline questionnaire includes a large set of sections describing the household structure (roster), the living conditions, education, assets, incomes, activities and spending. Some years, an occasional section is added to assess a specific issue.

The 2004 and 2005 rounds comprise all the baseline sections. However, one unique feature of the 2005 round is that it records more precisely the re-contact status of individuals in the sample. Indeed, for each resident member of the 2004 roster survey who was not re-contacted in 2005, enumerators were asked to register the reason why they were not able to re-interview him or why he left the household. One of these reasons was death of the member[5]. The 2004 round thus provides baseline socioeconomic information on households, while the 2005 round allows us to identify any adult death occurring between the two years. It is important to note that we also needed the 2006 round to compute the proper school attendance variables, because of the timing of the survey. Indeed, the fieldwork takes place at the end of the school year (between May and June) and the edu-

[4] The ROR was implemented as part of the MADIO project (Madagascar-DIAL-INSTAT-Orstom), established at the Malagasy National Institute of Statistics (INSTAT) and founded through the European Union (UE), the French "Institut de Recherche pour le Développement" (IRD, ex-Orstom) and the French Ministry of Cooperation. Depending on the donors' interest for such-and-such rural issue, some observatories were created, whereas other were phased out over time. There was a total of 13 observatories in 1999, 17 in 2000 and 2001 and 15 in 2002 and 2003. For more information, see the ROR webpage : http://www.dial.prd.fr/dial_enquetes/dial_enquetes_observatoires.htm.

[5] Other reasons were refusal, temporary absence, migration, marriage, divorce or back in the home household. The rate of missing values is 18% for this question. However, we are rather confident with the fact that death is well reported by the remaining members of the household, so that we do not underestimate the number of deaths in this setting.

cation section records school attendance for the previous year. Therefore, the 2005 round provides information on baseline children's attendance for the school year 2004-2005, i.e. before death occurs, whereas the 2006 round provides attendance status for the school year 2005-2006, i.e. after death. We have consequently a three year panel survey that allows us to estimate the impact of adult mortality only on two consecutive years.

Another unique feature of the 2004 survey is that it includes an occasional retrospective section collecting information on parental and matrimonial history for each adult. In particular, this section registers the occurrence and the time, in that event, of their parents' death, allowing us to evaluate the long-run impact of childhood orphanhood on completed schooling. We will turn to this point later in this paper (section 5).

3.2 Sample

Table 1 and 2 describe the sample structure, broken down into observatories, sex and orphan status. The 2004 baseline sample includes 3 616 households, that is 19 960 individuals, in 10 different observatories (33 villages)[6]. Enumerators managed to re-interview 82,2% of households and 84,1% of individuals in 2005 and 2006. Out of this sample, we were able to isolate a balanced panel of 6 095 school-aged children who constitute our unit of observation. A school-aged child is here defined as a child between 6 and 18. Indeed, primary school in Madagascar normally starts at 6 and comprises 6 grades. But secondary schooling is not unusual. That is why, unlike previous research, we extend this study to primary and secondary school enrollment. Among those children, 166 (2,7%) experienced an adult death in their household between the two rounds. They are referred to as "orphans".

Before processing further, two important comments have to be made. Firstly, death may appear as a marginal event in our setting. The reason is that, contrary to most studies focusing on the consequences of the HIV/AIDS epidemic, the ROR data we use

[6]The 2004 survey initially covered 14 observatories. Nevertheless, 4 observatories were phased out in 2005 and 2006. We assumed this "technical" attrition as random since it is related to lack of funding more than to specific village or household characteristics.

was not specifically designed to assess this issue. Besides, we have no information about the exact timing and the cause of the death. We thus consider in the analysis any adult death occurring between the two rounds, whatever its timing, its cause, the age and the link of the deceased member to surviving children, for the reasons explained in the previous section[7]. We may somewhat underestimate the number of deaths if failure to re-contact a whole household is due to the loss of one of its members (the head for instance). This brings us to our second point which is attrition.

The attrition rate in the children panel is 19,3%. If we add the 432 lost observations due to missing values on school attendance variables in the 2005 and 2006 rounds, the total number of observations excluded amounts to 25% of the sample. Despite a quite reasonable level, overall attrition is non negligible and may be an important issue for at least two reasons. First of all, attrition can be correlated with death, as mentioned above. More worryingly, attrition can also be correlated with our variable of interest, i.e. schooling. As the panel survey design is based on households and not on individuals, we have no schooling information neither on children from exited households nor on migrant children who left a re-interviewed household and were not traced. Basically, this paper estimates the immediate impact of adult mortality on schooling for those children remaining in the same (re-interviewed) household the year after death occurs. However, children out of the sample are likely to exhibit higher school enrollment if they moved to more favored areas, such as towns, or if they were fostered to wealthier households for instance[8]. The point is thus that attrition might induce some selection in our sample, regarding death and schooling, which might bias our results. We discuss this issue more precisely in the following section.

[7]However, 76% of the deaths recorded in the survey affect formerly resident parents of the sampled children, mostly in their working-age. The average age at death is indeed 56 years for men and 54 years for women

[8]This is one limitation of our study since were are not able to evaluate the potential mitigating effects of informal insurance networks through fostering and migration.

3.3 Preliminary Analysis

The choice of Madagascar in this paper is relevant with several respects. This unique island in the Indian ocean offers a broad ethnic and agroecological diversity which is well reflected in the ROR survey. Table 3 provides some descriptive statistics regarding the socioeconomic context of this study. In our sample, most rural households derive income from agricultural and pastoral activities, particularly rice growing. The average annual income per capita is 1 133 000 MGA (around 550 USD), while the average annual aggregated consumption per capita amounts to 785 000 MGA (around 380 USD). However, there are high disparities among areas and households. 7,4% of households consume less than the 1/2 median annual consumption per capita in this sample, that is less than 286 000 MGA (140 USD). Equally, Gini coefficients show that rice fields, which constitute the main households' assets, are unequally distributed. These figures reveal the high incidence of poverty and extreme deprivation in those rural areas, suggesting that households are likely to be vulnerable to shocks on their resources, such as death.

Nevertheless, Madagascar has been for long exemplary in terms of education outcomes within sub-Saharan Africa. Government's efforts to reduce tuition, improve access and foster partnerships with civil society made possible a considerable increase in primary gross enrollment, so that Madagascar may be one of the few developing countries achieving MDGs on education (at least a 100% primary gross enrollment rates by 2015). Secondary school enrollment is also not unusual. However, great disparities still exist within the countries, especially in rural regions, as displayed in table 3. Primary and secondary schooling remain lower than in urban areas and a great variability appears across observatories. Another unique feature of Madagascar is gender equity in education. Girls have now caught up with boys and sometimes display higher primary school enrollment (Cogneau et al., 2003). Investigating the impact of adult mortality in such a unique context is thus of great interest.

Turning now to this issue, table 4 shows school attendance transitions (between 2004 and 2005) of sampled children broken down by orphan status, that is before and after death occurs. A child is defined as attending school at period t if he was enrolled at the beginning of the school year and completed it. Two major findings emerge from table 4. Firstly, orphans are less likely to be enrolled in school both years. Besides, they are twice more likely to dropout the year after death. However, there is no significant difference on school entry. This table thus suggests that children are subjected to a greater risk of dropping out when experiencing death of an adult member in their household. But, orphans also seem to initially lay behind in terms of baseline schooling[9].

Table 5 displays a wider set of baseline summary statistics broken down by future orphan status. Simple tests of differences in means in the third column shows that orphans and non-orphans do not differ significantly along diverse measures of household wealth and schooling. This suggests that selection into orphanhood is not so strongly linked to initial socioeconomic discrepancies. Nevertheless, orphans tend to live in slightly larger households, to be older and less enrolled at baseline than their non-orphan counterparts. This must be kept in mind to further assess the robustness of our identification strategy to the extent that orphans present specific characteristics both correlated with death and especially schooling and which may confound the true impact of adult mortality.

Thus, if the previous figures give a first insight into the impact of adult mortality on school attendance, they do not account for a potential trend in schooling between the two years considered, the impact of other shocks affecting all children or effects of any other observable and unobservable child and household characteristics. To assess precisely differences in schooling trajectories between orphans and non-orphans and to make more profit of our longitudinal data set, we adopt in what follows a difference-in-differences identification strategy based on a well-defined counterfactual framework.

[9] Among reasons mentioned by respondents in orphans' households for not sending them to school in 2005, shortage of labour on crops and high school fees stand in first ans second positions, which is in line with our theoretical framework

4 Econometric Specification and Results

4.1 Identification Strategy

The identification strategy used in this paper relies on difference-in-differences methods (DID). Basically, it consists in likening adult death to a "treatment" (as in the randomized experimentation terminology) and to compare changes in schooling outcomes between the treatment group (T), that is orphans, and the control group (C), that is non-orphans, before and after death occurs. If we note D a dummy equal to 1 if a child becomes orphan in 2005 and S_{it} a schooling dummy equal to 1 if child i attend school at period t[10], the non-parametric DID estimator α of the impact of adult death on school attendance can be written as follows:

$$\alpha_{DID} = \mathbb{E}[S_{i1} - S_{i0}/D = 1] - \mathbb{E}[S_{i1} - S_{i0}/D = 0] \tag{1}$$

The experimental ideal would be to observe school attendance of orphans if they had not experienced death. However, we do not observe this potential outcome in the data, since a child cannot be in both the treatment and the control group. The fondamental assumption of the DID strategy is thus that orphans would have followed similar trend in school attendance than their non-orphans counterparts, had they not experienced an adult death in their household, so that differences in school attendance for non-orphans can be considered as the right conterfactual[11]. Under this identifying assumption, the DID estimator provides unbiased estimate of the average treatment effect, that is the average impact of death on schooling, controlling for any time-invariant household and child characteristics.

Table 6 provides this simple non-parametric DID estimate on the balanced panel of

[10] $t \in \{0, 1\}$ where 0 refers to the 2004 baseline period and 1 to the subsequent 2005 period, in what follows.

[11] Formally, if we denote S^0_{it} the potential schooling outcome of a child had he not experienced death, this identifying assumption can be written $\mathbb{E}[S^0_{i1} - S^0_{i0}/D = 1] = \mathbb{E}[S^0_{i1} - S^0_{i0}/D = 0] = \mathbb{E}[S_{i1} - S_{i0}/D = 0]$. It states that there is no selection into treatment, which will be further discuss in details.

school-aged children, broken down by gender. Results show that orphans school attendance drops by 8.8 percentage points between 2004 and 2005. The DID estimate shows that this decline is significantly different from the small decline in school attendance observed in the non-orphans control group by 6.8 percentage points. The effect is even more pronounced for female orphans.

However, this findings are only bivariate associations which may be spuriously driven by regional differences or household and child characteristics we want to simultaneously control for. We thus need to implement multivariate regressions which correspond to conditional versions of this benchmark non-parametric DID estimator.

4.2 Econometric Specification

Our estimation procedure follows three successive steps.

Schooling model with baseline controls: Considering that baseline differences between orphans and non-orphans exist and may be correlated with death and school attendance, we first estimate a flexible schooling model, with a wide set of baseline controls, including initial school attendance to capture the effect of initial observable heterogeneity in schooling outcomes:

$$S_{ih1} = \delta + \alpha D_{ih1} + \rho S_{ih0} + \beta X_{i0} + \gamma X_{h0} + \lambda V + \varepsilon_i \tag{2}$$

where S_{iht} is a dummy equal to 1 if child i in household h attends school at period t, D_{ih1} is a dummy equal to 1 if child i in household h becomes orphan at period 1, X_{i0} and X_{h0} are sets of baseline child and household characteristics, V is a whole set of village fixed-effects and ε_i is an individual error term allowed to be correlated within households. Note that we only control for 2004 baseline characteristics, since contemporaneous 2005 characteristics, such as household wealth or composition for instance, are likely to be affected by death itself and may confound its full impact. Equation (2) is close to a DID specification with a set of controls. To the extent that death can be considered as a random event conditional on those observable baseline characteristics, equation (2) yields an unbiased estimate α of the average impact of death on schooling.

DID schooling model with household fixed-effects: Despite the wide set of controls included in the previous equation, death may still not be considered as random and estimates may suffer from omitted variable bias, due to household and child unobservables, correlated with adult mortality and schooling, we want to equally control for. We can think, for instance, of specific behaviors toward health or households' involvement and preferences with respect to children schooling. We therefore estimate a second DID schooling model with household fixed-effects to take into account this additional unobserved heterogeneity:

$$S_{iht} = \delta_t + \eta_h + \alpha D_{iht} + \beta X_{i0} + \varepsilon_{it} \tag{3}$$

where δ_t is a time fixed-effect, η_h is a household fixed-effect and other variables are defined as above. Equation (3) is a regression equivalent of the simple non-parametric DID estimator in equation 1, conditional on observable child characteristics. Household fixed-effect controls for any unobservable time-invariant household characteristics, potentially correlated with adult death, that affect schooling, while capturing the effect of baseline household characteristics.

DID schooling model, with child fixed-effects: Similarly, considering that additional child unobservables, potentially correlated with death and schooling, may bias the last results, we finally estimate a DID schooling model with child fixed-effects:

$$S_{iht} = \delta_t + \eta_i + \alpha D_{ht} + \varepsilon_{it} \tag{4}$$

where η_i is a child fixed-effect and other variables are defined as above. Child fixed-effect captures the effect of any unobservable time-invariant child characteristics that affect schooling, as well as baseline household and child characteristics. Equation (4) is a regression equivalent of the DID estimator in section 4.1, conditional on individual time-invariant factors. To the extent that the unobserved differences between children who become orphans and those who do not are time-invariant, then equations (3) and (4) yield unbiased estimates α of the average impact of death on school attendance[12].

4.3 Results

Table 7 provides regression results from equation (2), (3) and (4). The schooling dependent variable S_{iht} is a dummy for school attendance. Note that a child is defined as attending school at period t if he was enrolled at the beginning of the school year and completed it. The independent variable of interest is a death dummy D_{iht} equal to 1 if child i in household h becomes orphan at period t[13]. The set of baseline child characteristics X_{i0} includes sex, age, age squared, and a dummy equal to 1 if child i is not a biological child of the household head[14]. The set of baseline household characteristics X_{h0} includes household composition by age and gender, sex and education of the household head (a dummy equal to 1 if the household head went to school), aggregated consumption, land assets (rice fields area and number of hill plots) and wealth/welfare indicators (number of rooms in the house, equipment index and distance to water in minutes). Unlike conventional DID specifications, we interact the time fixed-effect δ_t with a full set of village dummies in equation (3) and (4) to control for any specific community characteristics related to death and schooling (facilities or access to school, markets and various services for instance).

Finally, we estimate those three models with both Linear Probability Models (LPM) and Conditional Logit, with robust/clustered standards errors at the household level. Indeed, though Logit models are more suited to binary dependent variables, identification in Conditional Logit only relies on observations which exhibit variation regarding the

[12]Again, these results rely on the same fundamental DID identifying assumption which states that, conditional on observables and/or unobservables, there is no selection into treatment, i.e. death is as good as random. In this regression setting, this assumption can be formally re-written $P(D_{iht} = 1/\varepsilon_{it}) = P(D_{iht} = 1)$.

[13]In our setting, $D_{ih0} = 0$ for everyone, while D_{h1} switches to 1 if household h incurs an adult death between the two periods.

[14]In order to control for the potential inequality of treatment between household heads' own children and other child relatives with respect to schooling.

dependent variable. Thus, we also estimate LPM models to investigate the robustness of effects on the whole sample.

Results in table 7 suggest that adult mortality has an overall negative impact on subsequent children schooling decisions. The coefficient on the adult death dummy is significantly negative in all specifications and corresponds to a drop of about 20% in the probability to attend school for children incurring death in their household[15]. Estimated coefficients from LPM specifications are quite close to non-parametric DID estimates from table 6, which is not surprising since they correspond to conditional parametric versions of this benchmark estimator. What is more remarkable is that results are robust across all specifications, even after purging the potential correlation between adult death and unobservable time-invariant household and child characteristics. The average effect is even more pronounced in the last two specifications with child fixed-effects, though less precisely estimated[16]. Note also that the baseline school attendance dummy is strongly significant in the first two specifications, suggesting that failing to control for baseline schooling outcomes might severely bias the results in cross-sectional studies and that resorting to longitudinal data is crucial to consistently estimate the impact of adult mortality.

Other results are in line with well-known stylized facts regarding the demand for schooling in developing countries. Indeed, wealth indicators and household head schooling are positively correlated with school attendance. Estimates from the household fixed-effects models also show that older children and girls are more likely to attend school, whereas non-biological children of the household head seem to be disfavored. These latter results are consistent features of the Malagasy context, where first primary school enrollment is commonly observed between age 8 and 10 and girls are relatively favored

[15]Marginal change in probability calculated from logit specification (2). Unfortunately, deriving marginal effects from conditional logit models with fixed-effects remains quite tricky. Nevertheless, as the estimated coefficients are quite close in all specifications, we are confident with the fact that marginal changes in probability are of similar magnitude

[16]This is partly due to the fact that child fixed-effects models are "over-identified", since adult death shocks are measured at the household level. Thus, household fixed-effects probably capture most of the unobservable heterogeneity correlated with death

in terms of education. The result on non-biological children might suggest that fostering has no positive impact on schooling outcomes. However, we lack precise information in the data on those children foster (and maybe former orphan) status to credibly support this hypothesis. Finally note that specifications (1) and (2) do not exhibit such results regarding specific child characteristics, which is probably due to the fact that the baseline school attendance dummy captures most of these covariates effect.

5 Robustness checks

Though the previous estimations show a strong and robust adverse impact of adult mortality on schooling decisions, controlling for a set of various baseline characteristics as well as time-invariant household and child unobservables, we further test the validity of our results in this section, especially with respect to selection into treatment issues and attrition bias.

5.1 Selection into treatment issues

Indeed, one first major concern of DID estimates is its strong identifying assumption, which states that school attendance trajectories of non-orphaned sampled children are the right counterfactual. This may not hold if there is selection into treatment, that is to say if we suspect that orphans would have not followed parallel non-orphaned paths in school attendance, had they not incurred adult death in their household. This may be plausible if pre-treatment characteristics that are thought to be associated with the dynamics of the outcome are unbalanced between orphans and non-orphans. However, descriptive statistics from table 5 show that those two groups of school-aged children are quite similar along a wide set of baseline characteristics, except from age and initial household size. To further test the validity of the DID identifying assumption, the ideal would be to test the parallel trend hypothesis two periods before death occurred, i.e. between 2003 and 2004. Unfortunately, 2003 school attendance for sampled children is not available in the data, so that we cannot implement this "falsification" test. Yet, we can still estimate the impact of adult death one year preceding its occurrence. Concretely, we estimate the impact of a "future" death ($t = 1$) on baseline school attendance ($t = 0$), following this equation:

$$S_{ih0} = \delta + \alpha D_{ih1} + \beta X_{i0} + \gamma X_{h0} + \lambda V + \varepsilon_i \tag{5}$$

where S_{ih0} is a dummy equal to 1 if child i in household h attends school in 2004, D_{ih1} is a dummy equal to 1 if child i in household h becomes orphan in 2005 and other variables are defined as above. Table 8 gives estimates results from equation (5). They show that there is no significant difference between orphans and no-orphans regarding baseline school attendance, after controlling for baseline observable household and child characteristics. Formally, we test the following hypothesis : $\mathbb{E}[S_{i0}/D = 1, X] = \mathbb{E}[S_{i0}/D = 0, X]$, which states that the DID assumption collapses to a selection on observables restriction, which we accommodate in our estimations to estimate the average impact of adult mortality. One other interesting feature of the latter result is that the adverse impact of adult death does not seem to pre-date its occurrence, as found in previous studies (Yamano & Jayne, 2005; Evans & Miguel, 2007).

Nevertheless, this selection on observables restriction implies that all factors which confound simple comparisons of schooling outcomes between orphans and non-orphans are observed. This is a too stringent assumption if the distribution of unobserved variables is also believed to differ between this two groups. This would be the case for instance if selection into treatment is influenced by individual transitory shocks on past outcomes (Ashenfelter's dip) which creates non-parallel outcome dynamics, despite baseline observed similarities. Therefore, to finally assess the robustness of our previous results, we compute semi-parametric DID estimates, following Abadie (2005). He proposes a simple two-step weighting scheme based on the propensity score $P(D = 1/X)$ which is the only function which needs to be estimated in a first step. This weighted DID estimator writes:

$$\mathbb{E}\left[\frac{S_{i1}-S_{i0}}{P(D=1)}.\frac{D-P(D=1/X)}{1-P(D=1/X)}\right] \tag{6}$$

where S_{it} is a dummy equal to 1 if child i attends school at period t, D is the adult death dummy and X is a vector of household and child characteristics. Concretely, to estimate the average effect of death on school attendance, we thus need to estimate the propensity score and plug the fitted values into the sample analog of equation (6). This estimator is then robust to selection into treatment, provided that at least one fraction of the population is exposed to the treatment. Intuitively, this scheme works by weighting down the temporal difference in the outcome variable for non-orphans for those values of covariates which are over-represented among them and weighting-up this difference for those values of covariates under-represented[17]. Results are given in table 9 and show that the decline in school attendance for orphans is significantly different from the weighted decline in school attendance observed in the non-orphans control group by 6.2 percentage points. Though slightly smaller and less precise, this DID estimate is quite close to our previous results, lending further credence to the conclusion that selection into treatment does not strongly affects our results.

5.2 Attrition issues

Besides selection into treatment issues, one other selection bias concern is linked to attrition. As mentioned in section 3.2, among the initial 2004 sample of school-aged children, 19.3% were not reinterviewed in 2005 and 2006, either because their whole household was not recontacted (13.5% of observations) or because they left a recontacted household, mainly for migration, marriage or death reasons (5.8% of observations). If we take into account the 5.7% rate of missing values on the school attendance variable in the 2006 round, 25% of baseline observations were finally excluded from the analysis.

[17]This scheme builds on propensity score matching methods proposed by Heckman et al. (1998).

This non-negligible attrition might induce selection bias in our estimations for at least three reasons. Firstly, attrition may be correlated with death if failure to recontact a household is due to the death of its members, especially the household head[18]. Secondly, attrition may be correlated with schooling outcomes. Indeed, the main reason for attrition is migration, either of the whole household or of specific children within reinterviewed households. However, we do not have any schooling information on those exited children since they were not traced. Yet, they are likely to exhibit differential schooling outcomes if they moved to more or less favored areas or were fostered to more or less wealthy households for instance. Thirdly, non-response on school attendance variables might be related to specific characteristics of the respondents, such as illiteracy, directly correlated with children schooling. On the whole, the latter remarks suggest some potential issue of non-random attrition, which needs to be addressed in that event.

To investigate this issue, we first look at the determinants of attrition through a probit model where the dependant variable is a dummy equal to 1 if children were reinterviewed and school attendance was reported in 2005 and 2006. Covariates are the death dummy and baseline household and child characteristics. Results are given in table 10, column 1. They show that attrition is significatively correlated to death[19] as well as to several child and household characteristics such as age, gender, household composition and wealth. This suggest that attrition is a non-random event in our setting. To further assess the existence of an attrition bias, we use the approached suggested by Becketti et al. (1988) which consists in exploring the correlation between attrition and baseline outcome. To do so, we first regress 2004 children school attendance on baseline household and child characteristics plus a dummy for non-attrition, and secondly run the same regression adding a full set of non-attrition x covariates interaction terms. Table 11 gives the results. They show that not only is non-attrition significantly positively correlated with baseline school attendance, but also that the slope of coefficients in the latter regressions significantly

[18]However, only 0.9% of households were not reinterviewed for such a reason.

[19]This results stems from the fact that death of household members is better reported in reinterviewed households.

differs between attritors and non-attritors[20]. In other ways, attrition induces selection bias in our balanced panel, especially regarding our schooling outcome of interest.

To address this selection issue, we re-estimate the schooling models in section 4.2 corrected for attrition, using Heckman (1979) two-step procedure. Indeed, our empirical framework is comparable to a model with a latent attrition selection equation. We thus estimate in a first step the inverse probability of attrition, using as excluded instruments dummies for fieldwork surveyors. Our intuition is that, due to a great variability of recontact rates among surveyors, their identity is a strong determinant of attrition, while not being correlated to children schooling outcomes. We compute fitted values from this first equation, then calculate the inverse Mills ratio (IMR) that we plug as a correction term in the second stage schooling equations. The first stage equation corresponds to column 2 of table 10. The joint significance test of surveyor dummies at the bottom of the table shows that they are strong determinant of attrition and can be considered as valid instruments. Second stage estimates are reported in table 12. They show that attrition leads to an over-estimation of the impact of adult mortality. However, corrected coefficients are quite close to uncorrected ones and the IMR correction term is only weakly significant in the first two specifications[21]. On the whole, these findings suggest that attrition bias is quite small in our setting and probably strongly linked to time-invariant household and child unobservables we control for in fixed-effects models, so that we remain rather confident with our last specification results.

6 Heterogeneous effects and long-run persistence

If the previous results provide strong evidence that adult mortality adversely affects subsequent children schooling decisions, two questions have now to be raised. On the one hand, is the impact homogeneous across households and children ? The answer is of

[20]Formally, we test the Missing At Random (MAR) assumption (Little & Rubin, 1987), $\mathbb{E}[S_{i0}/X] = \mathbb{E}[S_{i0}/X, A_i = 1]$, where A_i is a dummy for non-attrition. The MAR assumption is rejected in our data.

[21]Note that we only present corrected LPM, to keep the sample size fixed but also because Heckman procedure is better suited to linear models

double interest. Firstly, it seems crucial for policy intervention to explore more accurately the differential vulnerability of individuals to unexpected death shocks. Secondly, the answer to this question will give insights into the underlying pathways through which death affects schooling. On the other hand, is their persistence in the long-run ? The issue here is to determine if schooling adjustments are transitory or if orphanhood matters in the long-run, in terms of total human capital accumulation.

6.1 Heterogeneity of effects

To investigate the heterogeneity of effects across households and children, we simply estimate diverse or "augmented" specifications of the DID schooling model with child fixed-effects (equation (4)). We first disaggregate the adult death dummy into two male and female death dummies to explore a differential impact according to the gender of the deceased household member. We then estimate alternative specifications including as additional regressors interaction terms between the adult death dummy and specific household and child characteristics such as gender, age and baseline index of poverty. Since inference based on estimated coefficients on interaction terms in Conditionnal Logit models is not direct[22], we only report LPM estimations.

Table 13 displays the results. The first finding is that the negative impact of adult mortality on schooling is mostly driven by male deaths, since only the male death dummy is significant in specification (1). In addition, girls, young orphans and orphans from poor households seems to share a larger part of the burden. These specific results are quite instructive about the underlying mechanisms at work. Indeed, as mentioned in section 2.1, the greater vulnerability of poor households, in line with the stronger effect of male mortality, support the hypothesis that the impact of death mainly passes through an income effect. After the death of one of its adult member, it seems likely that those afflicted household lack sufficient financial resources to continue children schooling and

[22] Ai & Norton (2003) show that interaction effects in nonlinear models cannot be evaluated simply by looking at the sign, magnitude or statistical significance of interaction term coefficients, estimated by standard softwares. It requires more sophisticated computation of "true" marginal effects.

maybe resort to child labor to compensate labor force shortage. Nevertheless, the fact that girl and young orphans are more deeply affected also indicates that expected returns to schooling are probably taken into account in the household decisions. Indeed, gender difference in returns to schooling is generally in favor of boys. As for the age effect, it appears relatively less costly to delay young children enrollment than to discontinue older ones schooling. However, these are only weak evidences. The size of our sample is too small to further investigate these specific issues.

6.2 Long run persistence

The previous analysis focused on the short run impact of adult mortality on subsequent children schooling decisions. Results suggest that household may resort to schooling adjustments as an immediate risk coping mechanism. Though, the decline in school attendance identified the year after death occurred might precisely correspond to transitory adjustments. The remaining issue is to explore the persistence of this adverse effect on human capital accumulation in the long run. This is a quite unstudied question in previous literature, basically due to the lack of long scope data. The only related paper by Beegle et al. (2010) on a long ten-year panel of Tanzanian children find that maternal orphanhood has a permanent adverse impact of one year of educational attainment.

In this prospect, we make use of an occasional retrospective section of the 2004 survey round which recorded, for each household head and spouse, information on their parents life history. Information was then collected on their place of residence, actual or former occupation, number of children, level of education, land and financial assets. One other particular feature of this section is that respondents were also asked to report parental death and, in that event, its exact timing. On the basis of this sample of adults with completed schooling and known orphan history, we are able to investigate the impact of childhood orphanhood on achieved levels of education[23].

[23] Note that the scope of the analysis is here restricted to the impact of parental death, since other adult deaths which occurred in the household during childhood were not recorded.

Nevertheless, since answers sometimes comes from old memories, one caveat concerning this retrospective section is the likelihood of misreported information, which raises the broader issue of measurement error. For instance, nearly half of orphaned respondents were not able to recall the exact timing of their parents death. In this case, they were asked to situate it in one of the three following age-groups : before age 5, between age 5 and 18 or after age 18. Table 14 describes the sample of 12 477 adults. About 30% of paternal (resp. maternal) orphans lost their father (resp. their mother) before age 18. On the whole, nearly one quarter of individuals in the sample lost at least one parent before age 18, i.e. during school-age. In what follows, this last category of individuals is referred to as (early) orphans, that is to say the "treatment" group. The "control" group is thus composed of individuals who did not loose any parent or lost one of them after age 18. Table 15 displays differences in schooling outcomes between treated and controls. Results show that individuals who incurred parental death during childhood are more likely to never have attended school than their non-orphaned counterparts. Indeed, average school enrollment, that is the proportion of children who were enrolled at least one time, among early orphans (33.7%) is significantly lower than school enrollment among non-orphans by 6.8 percentage points. Moreover, early orphans achieved in average about one year of education less. However, this basic bivariate differences are likely to be biased by inconsistent comparison of individual from different age-groups. Besides, they do not control for observable individual and parental characteristics.

To investigate more accurately the impact of early parental death on achieved level of education, we estimate the following long run schooling model :

$$E_i = \delta + \alpha D_i + \beta X_i + \gamma X_p + \lambda V + \varepsilon_i \tag{7}$$

where E_i is the number of completed years of schooling of individual i, D_i is a dummy equal to 1 if individual i lost one of his parents before 18, X_i is a set of individual characteristics including sex, date of birth, number of siblings and rank in the brotherhood,

X_p is a set of individual i's parents characteristics including levels of education and land assets, V is a whole set of origin region dummies and ε_i is an individual error term. As the dependent variable is left-censored since some individuals were never enrolled in school, we run Tobit regressions. Results are given in table 16. Specification (1) and (2) show that early orphanhood is strongly negatively correlated to the achieved level of education. Controlling for individual and parents characteristics, we find that early orphans completed in average 24% years of education less than their non-orphaned counterparts. This negative effect is ever more pronounced for individual who incurred a parental death before age 5 (specification (3)). This is due to the fact that very young orphans are more likely to have never been enrolled.

Though instructive, the latter regression does not make full use of available data. Indeed, for at least one part of the sampled individuals, we know the exact timing of parental deaths, as well as the year when they finished school. We thus estimate the following long run schooling model :

$$AgeE_i = \delta + \alpha D_i + \rho AgeD_i + \beta X_i + \gamma X_p + \lambda V + \varepsilon_i \tag{8}$$

where $AgeE_i$ is individual i's age when he finished schooling, $AgeD_i$ is his exact age when he incurred parental death and other variables are defined as above. Intuitively, equation (8) is close to a DID specification. Indeed, while the early orphanhood dummy D_i captures the effect of the treatment, the continuous variable $AgeD_i$ captures the effect of the timing of treatment. Identification lies on the fact that, among early orphans who incurred parental death during their childhood, average achieved level of education will be even lower if death occurred earlier. We estimate this equation on the sample of individuals who reported the exact timing of their parents death. As misreporting of information induces selection in our sample, potentially correlated with schooling outcomes[24], we correct estimations following Heckman approach, using surveyors dummies as instruments

[24]Indeed, individuals who correctly reported dates were younger and more educated in average.

in the first stage[25] (see section 5.2). Results are given in table 17. They show that the negative impact of parental death is even stronger when death occurred at young ages. These findings lend further support to the hypothesis that childhood orphanhood matters in the long run in terms of human capital accumulation.

7 Conclusion

The aim of this paper is to explore the link between adult death and children schooling decisions in rural Madagascar. We provide strong evidence that adult mortality has a negative immediate impact on afflicted children school attendance. Orphans are 20% less likely to attend school the year following death than their non-orphaned counterparts. The effect is even more stronger among girls, young orphans and children from relatively poorer households. If we cannot draw clear conclusions on the underlying mechanisms at work, the latter findings lend support to the assumption that this adverse effect is mainly driven by an income effect. These results suggest that, in a context where resources are scarce and formal insurance and market mechanisms are failing, households suffering unexpected shocks, such as death of one of its member, resort to children schooling adjustments and probably child labor as an immediate risk coping strategy. Pushing further the analysis to a sample of adults, we also provide evidence that adversity has long-lasting effect on human capital accumulation.

However, this last conclusion on the persistence of effects is clearly limited by the nature of our data. There is thus an urging demand for longitudinal data collection, in order to assess more accurately the impact of raising adult mortality rates due to HIV/AIDS on human capital accumulation and growth potential of developing economies.

[25] Concretely, we estimate in a first stage the probability to correctly report the timing of parental deaths. We thus compute fitted values and plug the Inverse Mills Ratio as a correction term in equation (8).

References

A. Abadie (2005). 'Semiparametric difference-in-differences estimators'. Review of Economic Studies **72**(1):1–19.

C. Ai & E. Norton (2003). 'Interaction terms in logit and probit models'. Economics Letters **80**(1):123–129.

M. Ainsworth, et al. (2005). 'The impact of adult mortality and parental deaths on primary schooling in North-Western Tanzania'. Journal of Development Studies **41**(3):412–439.

M. Ainsworth & D. Filmer (2002). Poverty, AIDS, and childrens schooling: a targeting dilemma. World Bank Development Research Group.

R. Akresh (2007). 'School enrollment impacts of non-traditional household structure'. Unpublished manuscript, University of Illinois at Urbana–Champagne. Retrieved July .

W. Bank (2002). 'Education and HIV/AIDs: A Window of Hope'. Washington DC, The World Bank .

G. Becker & N. Tomes (1979). 'An equilibrium theory of the distribution of income and intergenerational mobility'. The Journal of Political Economy **87**(6):1153–1189.

S. Becketti, et al. (1988). 'The Panel Study of Income Dynamics after fourteen years: an evaluation'. Journal of Labor Economics **6**(4):472–492.

K. Beegle (2003). 'Labor effects of adult mortality in Tanzanian households'. World .

K. Beegle, et al. (2006). Adult mortality and consumption growth in the age of HIV/AIDS. World Bank, Development Research Group, Poverty Team.

K. Beegle, et al. (2010). 'Orphanhood and human capital destruction: Is there persistence into adulthood?'. Demography **47**(1):163–180.

G. Bicego, et al. (2003). 'Dimensions of the emerging orphan crisis in sub-Saharan Africa'. Social Science & Medicine **56**(6):1235–1247.

A. Case & C. Ardington (2006). 'The impact of parental death on school outcomes: Longitudinal evidence from South Africa'. Demography **43**(3):401–420.

A. Case, et al. (2004). 'Orphans in Africa: parental death, poverty, and school enrollment'. Demography **41**(3):483–508.

D. Cogneau, et al. (2003). 'Madagascar'. Morrisson C.(ed.), Depenses publiques d education et de sante et reduction de la pauvrete en Afrique de l Est: Madagascar et Tanzanie, Etudes du Centre de Developpement de l OCDE, Paris pp. 17–113.

D. Evans & E. Miguel (2007). 'Orphans and schooling in Africa: A longitudinal analysis'. Demography **44**(1):35–57.

G. Foster, et al. (1995). 'Orphan prevalence and extended family care in a peri-urban community in Zimbabwe'. AIDS care **7**(1):3–18.

P. Gertler, et al. (2004). 'Schooling and parental death'. Review of Economics and Statistics **86**(1):211–225.

J. Heckman (1979). 'Sample selection bias as a specification error'. Econometrica: Journal of the econometric society pp. 153–161.

J. Heckman, et al. (1998). 'Matching as an econometric evaluation estimator'. Review of Economic studies **65**(2):261–294.

S. Hunter & J. Williamson (2000). 'Children on the Brink'. USAID, Washington DC .

A. Kamali, et al. (1996). 'The orphan problem: experience of a sub-Saharan Africa rural population in the AIDS epidemic'. AIDS care **8**(5):509–516.

R. Little & D. Rubin (1987). 'Statistical analysis with missing data' .

C. Lloyd & A. Blanc (1996). 'Children's schooling in sub-Saharan Africa: The role of fathers, mothers, and others'. Population and Development Review **22**(2):265–298.

M. McPherson (2001). 'HIV/AIDS impacts on capacity deepening and economic growth'. Brookings Institution, 28th June. Washington DC .

V. Naidu & G. Harris (2005). 'The impact of HIV/AIDS morbidity and mortality on households–a review of household studies'. South African Journal of Economics **73**(s1):533–544.

R. Ryder, et al. (1994). 'AIDS orphans in Kinshasa, Zaire: incidence and socioeconomic consequences'. Aids **8**(5):673.

R. Townsend (1994). 'Risk and insurance in village India'. Econometrica: Journal of the Econometric Society **62**(3):539–591.

R. Townsend (1995). 'Consumption insurance: an evaluation of risk-bearing systems in low-income economies'. The Journal of Economic Perspectives **9**(3):83–102.

U. UNAIDS (2002). 'USAID, Children on the Brink 2002: A Joint Report on Orphan Estimates and Program Strategies'.

T. Yamano & T. Jayne (2004). 'Measuring the impacts of working-age adult mortality on small-scale farm households in Kenya'. World Development **32**(1):91–119.

T. Yamano & T. Jayne (2005). 'Working-age adult mortality and primary school attendance in rural Kenya'. Economic Development and Cultural Change **53**(3):619–653.

Tables

Table 1: **Household and individual sample composition**

Observatories	**Baseline sample 2004**		**Attrition (%) 2005/2006**	
	Households	**Individuals**	**Households**	**Individuals**
Antsirabe	132	752	25.8	25.5
Marovoay	515	2 723	7.9	6.9
Farafangana	289	1 834	27.7	27.7
Ambovombe	421	2 448	16.1	7.2
Ambatondrazaka	503	2 675	3.9	3.9
Fenerive	100	450	8.9	9.1
Mahanoro	199	1 119	4.1	4.1
Morondava	485	2 444	48.4	48.1
Manandriana	471	2 512	17.6	17.6
Tsivory	501	3 003	12.8	12.8
Total	**3 616**	**19 960**	**17,8**	**15,9**

Source: ROR Surveys, 2004-2006

Table 2: **School-aged children sample composition**

	Baseline sample 2004	**Surveyed in 2005/2006**	**Attrition (%)**
Non-orphans	7 362	5 929	19.5
Boys	*3 880*	*3 183*	*18.0*
Girls	*3 482*	*2 746*	*21.1*
Orphans	189	166	12.2
Boys	*105*	*94*	*10.5*
Girls	*84*	*72*	*14.3*
Total	**7 551**	**6 095**	**19.3**

Source: ROR Surveys, 2004-2006

Table 3: **Wealth, inequality indexes and school enrollment (2004)**

Observatories	**Mean income per capita[(a)] (thds of mga)**	**Mean consumption per capita (thds of mga)**	**Poor households[(b)] (%)**	**Gini (rice fields)**	**School enrollment[(c)]**	
					Girls	**Boys**
Antsirabe	1 074	829	11.5	0.53	78.7	80.2
Marovoay	1 595	1 029	3.7	0.42	67.3	65.6
Farafangana	944	667	4.2	0.47	58.1	59.8
Ambovombe	564	482	11.4	(-)	79.4	56.3
Ambatondrazaka	1 308	788	5.2	0.53	77.2	71.8
Fenerive	889	742	1.2	0.46	83.6	80.6
Mahanoro	726	602	1.1	0.45	85.5	82.9
Morondava	1 295	846	9.7	0.42	55.4	48.1
Manandriana	933	727	4.5	0.45	78.9	75.4
Tsivory	1 258	765	15.6	0.45	50.6	38.6
Total	**1 133**	**785**	**7,4**	**0,51**	**69,8**	**61,8**

Source: ROR Surveys,2004-2006

Notes:

[(a)] Income and consumption are expressed in thousands of malagasy ariary (MGA). One thousand MGA equals 0.5 USD. Consumption is the aggregate of food and non-food expenditures.

[(b)] Poor households are defined as households consuming less than the 1/2 median annual consumption per capita in this sample.

[(c)] School enrollment rates refer to children aged 6-18.

Table 4: **School attendance transitions (2004-2005)**

	Non-orphans (%)	**Orphans** (%)
Never attend (No-No)	28.9	35.9
Dropout (Yes-No)	6.3	13.2
Entry (No-Yes)	4.3	4.4
Attend both years (Yes-Yes)	60.5	46.5
Observations	**5 063**	**159**

Source: ROR Surveys, 2004-2006

Note: Sample restricted to children aged 6-18 both years.

Table 5: **Household and child baseline characteristics by orphan status (2004)**

	Non-orphans (C)		Orphans (T)		Difference[(a)] $(T) - (C)$	
	mean	**sd**	**mean**	**sd**	**mean**	**sd**
Child characteristics						
School enrollment (d)[(b)]	0.67	0.47	0.60	0.49	-0.07**	0.04
Girl (d)	0.47	0.50	0.43	0.50	-0.04	0.04
Age	11.14	3.52	11.81	3.88	0.67**	0.05
Non-biological child (d)	0.13	0.34	0.16	0.37	0.03	0.03
Observations	**5 063**		**159**			
Household characteristics						
Size	6.45	2.35	7.05	2.69	0.60**	0.31
Number of adult men	1.19	0.74	1.49	0.90	0.30**	0.20
Number of adult women	1.23	0.59	1.54	0.84	0.32***	0.08
Number of children	4.02	1.86	4.03	1.96	0.01	0.09
Female headed (d)	0.18	0.38	0.17	0.38	-0.01	0.05
Head went to school (d)	0.70	0.46	0.64	0.48	-0.06	-0.06
Head can read (d)	0.72	0.45	0.69	0.46	-0.03	0.06
Head can write (d)	0.72	0.45	0.69	0.46	-0.03	0.06
Annual consumption (thds of mga)	4 182	3 327	3 722	2 109	-460	436
Landowner (d)	0.87	0.33	0.81	0.39	-0.04	0.05
Ricefield area (ares)	4.36	1.04	4.38	1.11	0.01	0.18
Hill plots (number)	1.30	0.57	1.30	0.63	0.00	0.09
Number of house rooms	1.99	1.18	1.98	1.20	0.00	0.16
Equipment index[(c)]	1.92	2.25	1.91	2.34	-0.01	0.27
Distance to water (min)	25.71	36.50	27.80	38.76	2.09	4.82
Observations	**2 035**		**98**			

Source: ROR Surveys, 2004-2006

Notes:

[(a)] *** 1% significance level, **5% significance level, from a test for differences between columns (T) and (C).

[(b)] The suffix (d) stands for dummy variables.

[(c)] The equipment index corresponds to the first component from a PCA on households goods, such as chairs, tables, beds, tv, radio sets, bicycles, motorcycles and cars.

Table 6: **Non-parametric difference-in-differences estimates**

School enrollment (%)	**Non-orphans** (C) **2004**	**2005**	**Orphans** (T) **2004**	**2005**	$\Delta(C)$	$\Delta(T)$	**DID** $\Delta(T)-\Delta(C)$
Whole sample	66.7	64.7	50.9	59.7	-0.2	-8.8	- 6.8***
	(0.01)	(0.01)	(0.04)	(0.04)	(0.01)	(0.03)	(0.03)
Boys	62.6	61.1	55.6	63.3	-1.5	-7.8	-6.3**
	(0.01)	(0.01)	(0.05)	(0.05)	(0.01)	(0.04)	(0.03)
Girls	71.5	68.9	44.9	55.1	-2.6	-10.1	-7.5**
	(0.01)	(0.01)	(0.06)	(0.06)	(0.01)	(0.06)	(0.04)
Observations	**5 063**		**159**				

Source: ROR Surveys, 2004-2006

Note: Standard errors in parenthesis.

*** 1% significance level, **5% significance level, issued from an univariate linear regression of $\Delta(S)$ on a death dummy D

Table 7: **Schooling models with death shocks, estimation results**

	With baseline controls		**DID with household fixed-effects**		**DID with child fixed-effects**	
	LPM (1)	**Logit (2)**	**LPM (3)**	**Logit (4)**	**LPM (5)**	**Logit (6)**
Adult death (d)	-0.066***	-0.711***	-0.049**	-0.681**	-0.069*	-0.755*
	(0.021)	(0.234)	(0.016)	(0.324)	(0.037)	(0.406)
Baseline school attendance (d)	0.664***	3.654***				
	(0.016)	(0.121)				
Child characteristics (2004)						
Age	0.004	0.099	0.186***	1.873***		
	(0.009)	(0.091)	(0.010)	(0.078)		
Age squared	-0.001***	-0.014***	-0.010***	-0.096***		
	(0.000)	(0.004)	(0.000)	(0.004)		
Girl	0.003	-0.018	0.031***	0.364***		
	(0.009)	(0.094)	(0.011)	(0.086)		
Non-biological child (d)	-0.014	-0.172	-0.091***	-0.704***		
	(0.013)	(0.142)	(0.027)	(0.168)		
Household characteristics (2004)						
Number of adult men	-0.013*	-0.142*				
	(0.008)	(0.079)				
Number of adult women	0.004	0.025				
	(0.010)	(0.111)				
Number of boys	-0.004	-0.050				
	(0.004)	(0.041)				
Number of girls	0.001	0.016				
	(0.004)	(0.039)				
Female headed (d)	-0.013	-0.117				
	(0.016)	(0.162)				
Head went to school (d)	0.025*	0.227*				
	(0.013)	(0.134)				
Annual consumption (thds of mga)	-0.005	-0.068				
	(0.010)	(0.113)				
Landowner (d)	0.015	0.148				
	(0.022)	(0.236)				
Ricefield area owned (ares)	0.004	0.045				
	(0.003)	(0.033)				
Number of hill plots owned	-0.002	-0.008				
	(0.009)	(0.109)				
Number of house rooms	0.004	0.056				
	(0.005)	(0.059)				
Equipment index	0.010***	0.125***				
	(0.002)	(0.032)				
Distance to water (min)	0.006	0.050				
	(0.006)	(0.066)				
Constant	0.406***	0.264				
	(0.151)	(1.648)				
Time x community dummies	yes	yes	yes	yes	yes	yes
Observations	5222	5222	10444	10444	10444	10444

Source: ROR Surveys, 2004-2006

Notes: Dependent variable is school attendance in 2005 in specifications (1) and (2), and school attendance at period t in specifications (3) to (6). Estimated coefficients are reported, with clustered/robust standard errors in parenthesis. *** indicates 1% significance level, ** 5% significance level and * 10% significance level. (d) stands for dummy variables. Sample is restricted to children aged 6-18 both years

Table 8: **Baseline schooling model with future death shocks, estimation results**

	LPM (1)	**Logit (2)**
"Future" adult death	-0.024	-0.117
	(0.048)	(0.375)
Constant	-0.819***	-9.523***
	(0.215)	(1.691)
Child characteristics	yes	yes
Household characteristics	yes	yes
Community dummies	yes	yes
Observations	5222	5222
R^2	0.37	0.36

Source: ROR Surveys, 2004-2006

Notes: Dependent variable is school attendance in 2004. Estimated coefficients are reported, with clustered/robust standard errors in parenthesis. *** indicates 1% significance level, ** 5% significance level and * 10% significance level. Sample is restricted to children aged 6-18 both years.

Table 9: **Semi-parametric difference-in-differences estimates**

	Non-orphans (weighted) $\Delta(C)$	**Orphans** $\Delta(T)$	**Semi-parametric DID** $\Delta(T) - \Delta(C)$
$(S_{i1} - S_{i0})$	-0.026	-0.088	-0.062*
	(0.010)	(0.030)	(0.032)
Observations	**5 063**	**159**	

Source: ROR Surveys, 2004-2006

Note: Results computed with Abadie's software package available online

Dependant variable is school attendance at period t. Covariates used in the first step estimation of the propensity score are the whole set of baseline household and child characteristics. Standard errors are in parenthesis, * indicates 10% significance level.

Table 10: **Determinants of attrition, probit models**

	Probit (1)	**Probit (2)**
Adult death (d)	0.131***	0.131***
	(0.022)	(0.041)
Child characteristics (2004)		
Age	0.041***	0.041***
	(0.008)	(0.009)
Age squared	-0.001***	-0.002***
	(0.000)	(0.000)
Girl	-0.036***	-0.036***
	(0.011)	(0.013)
Non-biological child (d)	0.162***	0.163**
	(0.012)	(0.021)
Household characteristics (2004)		
Number of adult men	-0.015*	-0.015
	(0.007)	(0.011)
Number of adult women	-0.010	-0.010
	(0.009)	(0.014)
Number of boys	0.009***	0.009*
	(0.004)	(0.004)
Number of girls	0.019***	0.019***
	(0.004)	(0.005)
Female headed (d)	-0.037*	-0.039*
	(0.017)	(0.018)
Head went to school (d)	-0.007	-0.007
	(0.013)	(0.021)
Annual consumption (thds of mga)	0.001	0.001
	(0.011)	(0.015)
Landowner (d)	0.026	0.030
	(0.025)	(0.031)
Ricefield area owned (ares)	-0.003	-0.004
	(0.003)	(0.004)
Number of hill plots owned	-0.001	-0.001
	(0.010)	(0.015)
Number of house rooms	0.002	0.002
	(0.006)	(0.062)
Equipment index	0.006**	0.005**
	(0.002)	(0.002)
Distance to water (min)	0.004	0.003
	(0.045)	(0.045)
Constant	-0.290	-0.418
	(0.818)	(0.731)
Community dummies	yes	yes
Surveyor dummies	no	yes
Observations	**7551**	**7551**
Test of joint significance of surveyors dummies (χ^2)	(-)	279.1*** (0.000)

Source: ROR Surveys, 2004-2006

Note: Dependant variable is a dummy equal to 1 if children were reinterviewed and school attendance was reported in 2005/2006 rounds. Coefficient are reported, clustered standard errors are in parenthesis. *** indicates 1% significance level, ** 5% significance level and * 10% significance level. (d) stands for dummy variables.

Table 11: **Test of attrition bias on baseline schooling outcome**

	Logit (1)	**Logit (2)**
Non-attrition dummy	0.219**	0.170*
	(0.089)	(0.094)
Child characteristics	yes	yes
Household characteristics	yes	yes
Community dummies	yes	yes
Non-attrition interaction terms	no	yes
Observations	**7551**	**7551**
Test of joint significance of interactions	(-)	149.43***
terms without the constant (χ^2)		(0.000)
Test of joint significance of interactions	(-)	63.33***
terms with the constant (χ^2)		(0.000)

Source: ROR Surveys, 2004-2006

Note: Dependant variable is a dummy equal to 1 if children attend school in 2004. Coefficient are reported, clustered standard errors are in parenthesis.*** indicates 1% significance level, ** 5% significance level and * 10% significance level.

Table 12: **Schooling models with death shocks, LPM corrected for attrition**

	With baseline controls (1)	**DID with household fixed-effects (2)**	**DID with child fixed-effects (3)**
Adult death (d)	-0.063**	-0.044**	-0.067**
	(0.031)	(0.021)	(0.039)
Inverse Mills ratio	-0.002*	-0.001*	-0.001
	(0.001)	(0.000)	(0.001)
Constant	-0.754***		
	(0.133)		
Baseline school attendance	yes	no	no
Child characteristics	yes	yes	yes
Household characteristics	yes	yes	yes
Time x community dummies	yes	yes	yes
Observations	**5 222**	**10 444**	**10 444**

Source: ROR Surveys, 2004-2006

Notes: Dependent variable is school attendance in 2005 in specifications (1), and school attendance at period t in specifications (2) and (3). Estimated coefficients are reported, with boostrap standard errors in parenthesis. *** indicates 1% significance level, ** 5% significance level and * 10% significance level. (d) stands for dummy variables. Sample is restricted to children aged 6-18 both years

Table 13: **Schooling models with child fixed-effects, heterogenous effects**

	LPM (1)	**LPM (2)**	**LPM (3)**	**LPM (4)**
Adult death		-0.066**	-0.200*	-0.067*
		(0.032)	(0.109)	(0.039)
Male adult death	-0.143***			
	(0.046)			
Female adult death	-0.005			
	(0.066)			
Adult death x Girl		-0.022**		
		(0.011)		
Adult death X Age			0.010*	
			(0.006)	
Adult death x Poor				-0.041*
				(0.024)
Time x community dummies	yes	yes	yes	yes
Observations	**10 444**	**10 444**	**10 444**	**10 444**

Source: ROR Surveys, 2004-2006

Notes: Dependent variable is school attendance at period t. Estimated coefficients are reported, with clustered standard errors in parenthesis. *** indicates 1% significance level, ** 5% significance level and * 10% significance level. Sample is restricted to children aged 6-18 both years. A household is defined as poor if he consumes less than the 1/2 median annual consumption per capita.

Table 14: **Adult sample composition and orphanhood prevalence**

	Paternal (%)	**Maternal (%)**
Orphans	**55.1**	**43.7**
Before age 5	9.9	9.4
Between age 5 and 18	22.3	19.2
After age 18	67.8	71.4
Non-orphans	**40.5**	**53.5**
Missing values	4.4	2.8
Observations	**12 477**	**12 477**

Source: ROR Surveys, 2004

Table 15: **School enrollment and average completed years of education, by early orphan status**

	Orphans (T)	**Non-orphans** (C)	**Difference** $(T)-(C)$
School enrollment (%)	**66.3**	**73.1**	**-6.8*** **
Male	70.9	76.8	-5.9***
Female	62.1	69.5	-7.4***
Years completed	**2.8**	**3.4**	**-0.6*** **
Male	3.2	3.8	-0.6***
Female	2.4	3.1	-0.7***
Observations	**2 998**	**8 929**	

Source: ROR Surveys, 2004

Notes: Early orphans are defined as individuals who lost at least one parent. before the age of 18. An individual is said to be enrolled in school if he was enrolled at least one time. Last column is a simple test of equality of means. *** indicates 1% significance level

Table 16: **Long run schooling model, Tobit regressions**

	Tobit (1)	**Tobit (2)**	**Tobit (3)**
Orphan before 18 (d)	-0.540***	-0.241***	
	(0.085)	(0.091)	
Orphan before 5 (d)			-0.346***
			(0.142)
Orphan between 5 and 18 (d)			-0.194***
			(0.106)
Individual characteristics			
Female (d)		-0.756***	-0.755***
		(0.078)	(0.078)
Date of birth		0.028***	0.028***
		(0.003)	(0.003)
Number of siblings		0.045***	0.045***
		(0.013)	(0.013)
First born (d)		-0.140	-0.142**
		(0.112)	(0.113)
Last born (d)		-0.174**	-0.174**
		(0.117)	(0.117)
Parents characteristics			
Father education level		0.181***	0.180***
		(0.020)	(0.020)
Mother education level		0.178***	0.181***
		(0.028)	(0.028)
Common ricefield area (ares)		0.050***	0.050***
		(0.024)	(0.024)
Father ricefield area (ares)		0.027*	0.029*
		(0.023)	(0.023)
Mother ricefield area (ares)		-0.019	-0.021
		(0.259)	(0.226)
Constant	6.266***	-75.080***	-74.245***
	(0.363)	(5.807)	(5.797)
Region dummies	yes	yes	yes
Observations	11765	9380	9380

Source: ROR Surveys, 2004

Notes: Dependant variable is completed years of schooling. Marginal effects reported, standard errors in parenthesis. *** indicates 1% significance level, ** 5% significance level and * 10% significance level. (d) stands for dummy variables.

Table 17: **Long run schooling model, OLS regressions**

	OLS	**Corrected OLS**
	(1)	**(2)**
Orphan before 18	-0.417***	-0.356***
	(0.086)	(0.076)
Age at death	0.053***	0.042***
	(0.010)	(0.012)
Inverse Mills ratio		-0.001*
		(0.000)
Constant	-16.336***	-15.344***
	(4.522)	(4.624)
Individual characteristics	yes	yes
Parents characteristics	yes	yes
Region dummies	yes	yes
Observations	7702	7702

Source: ROR Surveys, 2004

Notes: Dependant variable is age when schooling finished. Coefficients reported, robust/ bootstrap standard errors in parenthesis. *** indicates 1% significance level, ** 5% significance level and * 10% significance level.