



# The effect of maternal tetanus immunization on children's schooling attainment in Matlab, Bangladesh: Follow-up of a randomized trial

## Citation

Canning, David, Abdur Razzaque, Julia Driessen, Damian G. Walker, Peter Kim Streatfield, and Mohammad Yunus. 2011. The effect of maternal tetanus immunization on children's schooling attainment in Matlab, Bangladesh: Follow-up of a randomized trial. *Social Science & Medicine* 72, no. 9: 1429–1436. doi:10.1016/j.socscimed.2011.02.043.

## Published Version

doi:10.1016/j.socscimed.2011.02.043

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:33730174>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available. Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)



**PROGRAM ON THE GLOBAL  
DEMOGRAPHY OF AGING**

**Working Paper Series**

**The Effect of Maternal Tetanus Immunization on  
Children's Schooling Attainment in Matlab,  
Bangladesh: Follow-up of a Randomized Trial**

David Canning<sup>1</sup>, Abdur Razzaque<sup>2</sup>, Julia Driessen<sup>3</sup>,  
Damian G. Walker<sup>3</sup>, Peter Kim Streatfield<sup>2</sup>, Mohammad Yunus<sup>2</sup>

<sup>1</sup>Harvard University; <sup>2</sup>ICDDR,B; <sup>3</sup>Johns Hopkins University;

July 2010

PGDA Working Paper No. 76

<http://www.hsph.harvard.edu/pgda/working.htm>

The views expressed in this paper are those of the author(s) and not necessarily those of the Harvard Initiative for Global Health. The Program on the Global Demography of Aging receives funding from the National Institute on Aging, Grant No. 1 P30 AG024409-06.

# The Effect of Maternal Tetanus Immunization on Children's Schooling Attainment in Matlab, Bangladesh: Follow-up of a Randomized Trial

David Canning<sup>1</sup>, Abdur Razzaque<sup>2</sup>, Julia Driessen<sup>3</sup>,  
Damian G. Walker<sup>3</sup>, Peter Kim Streatfield<sup>2</sup>, Mohammad Yunus<sup>2</sup>

<sup>1</sup>Harvard University; <sup>2</sup>ICDDR,B; <sup>3</sup>Johns Hopkins University;

July 2010

## Abstract

We investigate the effects of ante-natal maternal vaccination against tetanus on the schooling attained by children in Bangladesh. Maternal vaccination prevents the child from acquiring tetanus at birth through blood infection and substantially reduces infant mortality and may prevent impairment in children who would otherwise acquire tetanus but survive. We follow up on a 1974 randomized trial of maternal tetanus toxoid, looking at outcomes for children born in the period 1975-1979. We find significant schooling gains from maternal tetanus vaccination for children whose parents had no schooling, showing a large impact on a small number of children.

**Key words:** Vaccination, tetanus, schooling, education, Bangladesh.

-----  
Name of corresponding author: David Canning, Harvard School of Public Health.  
Email: dcanning@hsph.harvard.edu

## **Introduction**

There is a growing literature that links interventions in early childhood health and nutrition to improved educational attainment (Bleakley 2003; Miguel and Kremer 2004; Bleakley 2006; Jukes, Pinder et al. 2006; Maluccio, Hoddinott et al. 2009; Bleakley 2010; Cutler, Fung et al. 2010; Lucas 2010) and earnings as an adult (Hoddinott, Maluccio et al. 2008; Bleakley 2010). We add to this evidence base by examining the outcomes of a randomized trial to examine the effect of protection against tetanus in early childhood on years of schooling attained.

Previous analyses of the benefits of vaccines have found large effects on infant mortality (Koenig 1992; Breiman 2004). It has been proposed that the vaccination of children also reduces their morbidity, and aids their physical and cognitive development, resulting in long run improvements in educational outcomes, and eventually in adult productivity and earnings (Bloom, Canning et al. 2005). If this is true, vaccination can be seen not only as a health intervention that lowers mortality but as an economic intervention that has economic returns in terms of worker productivity and income levels (Bloom and Canning 2000). While it seems reasonable that lower mortality and reduced morbidity go hand in hand, and that early childhood exposures to infectious disease can have long run effects on adults, the direct evidence base for an effect of vaccination on education, and eventually earnings, is weak. We address the issue in this paper by following up children born after a randomized trial of a tetanus toxoid vaccine administered to women in Matlab, Bangladesh in 1974 and looking at the effect of the treatment on their children's schooling attainment in 1996.

The main cause of tetanus is through contamination when the child's umbilical cord is cut at birth. Tetanus spores are ubiquitous in soil and can be harbored on skin or rusty metals. Cutting an infant's umbilical cord, and dressing the wound, creates opportunities for infection if high standards of cleanliness are not maintained. Maternal vaccination against tetanus, however,

creates antibodies that are passed on to the child in utero, giving protection to the disease in newborns. We examine outcomes for children born between 1975 and 1979 to mothers who participated in the 1974 study, comparing children whose mothers received the tetanus toxoid vaccine with those whose mothers did not. Our studies adds to the existing evidence for intergenerational transmission of health effects (Behrman, Calderon et al. 2009; Subramanian, Ackerson et al. 2009).

The effects of maternal vaccination in this trial on infant mortality have already been examined (Black, Huber et al. 1980; Koenig, Roy et al. 1998). There was a substantial reduction in mortality in the first month of life (neonatal mortality) among children of women who received the tetanus toxoid, with a decline in the neonatal mortality rate from around 70 per 1000 to around 40 per 1000.

Neonatal tetanus has a high case fatality rate – between 30% and 70% of infants admitted to hospitals in developing countries with neonatal tetanus die of the disease (Brenzel, Wolfson et al. 2006). Among those children who acquire neonatal tetanus and survive, perhaps 20% have substantial disability while up to 50% may suffer some cognitive impairment (Roper, Vandelaer et al. 2007). Without immunization, a number of children will acquire neonatal tetanus and die while some will acquire the infection and survive, but with impairments that may prevent educational success. We expect to see the effects of this cognitive impairment in schooling outcomes.

There is evidence of a socio-economic gradient in the effects of ill health on children. The effect of infectious disease is often more pronounced in children from families of low economic status, since the ability to combat infection is linked to nutritional status and access to health care (Koenig, Bishai et al. 2001; Black, Morris et al. 2003). We therefore examine if the

effect of maternal vaccination on children's schooling outcomes varies with the parental level of schooling, which we use as a proxy for the family's socioeconomic status.

While we have a randomized trial, a major problem with our data is that by 1996 about a third of our sample of children had migrated out of the Matlab area and we do not have follow-up measurement of their school attainment. We address this issue by modeling the migration decision as a function of schooling attainment, taking migration just after 1996 as depending on the 1996 level of schooling attained. This allows us to correct our results for induced migration before 1996 on the assumption it followed the same pattern. School attainment is a significant predictor of outmigration, though correcting our estimates for this selective outmigration has little impact on the estimated effect of the tetanus intervention.

We find a significant positive effect of maternal tetanus vaccination on the school attainment of children from families with low socioeconomic status (whose parents themselves have zero years of schooling). About 3% of children from zero schooling parents are moved from no schooling attainment to achieving 8 or more years of schooling if their mothers received the maternal tetanus vaccination. This concentrated large effect in a small number of children is consistent with the idea that tetanus infections are rare, but can cause substantial impairment when they occur. We estimate an average gain of around 0.25 years of schooling for children whose parents had no schooling when their mothers are vaccinated against tetanus.

Neonatal tetanus is now rare in developed countries but it still generates a high disease burden in Africa and Asia (World Health Organization 2006). Neonatal tetanus mortality rates of between 23 and 82 per 1000 births are common in un-immunized populations, and large reductions in neonatal mortality have been observed worldwide as maternal tetanus immunization has been rolled out (Meegan, Conroy et al. 2001; Roper, Vandelaer et al. 2007).

As well as seeing reductions in mortality due to this rollout our results from Matlab suggest we should also see improved schooling outcomes. A simple cost benefit analysis suggests that maternal vaccination against tetanus is a highly cost effective method of increasing school attainment when compared with other mechanisms. Adding these educational gains to the direct mortality improvements from maternal vaccination adds to the case for making the vaccination of pregnant women against tetanus a priority.

### **The Intervention and Data**

In July – August 1974, non-pregnant women over 15 years old in Matlab were enrolled in a double blind clinical trial and were randomly assigned to either cholera vaccine or tetanus-diphtheria toxoid vaccine. Women received at least one dose and about three quarters of women received a follow up second dose 42 days later. Each dose was administered by injection. The cholera vaccine, the focus of the study, was ineffective, providing only about three months of protection against cholera to about 40% of recipients (Curlin, Levine et al. 1978). In addition, the diphtheria part of the tetanus-diphtheria vaccine was unlikely to have had any effect on the women because pre-immunization testing indicated that everyone in the population over six years of age already had immunity to diphtheria (Feeley, Curlin et al. 1979). However, the tetanus toxoid part of the vaccine has been documented to be effective at reducing neonatal mortality in the children born to these women in subsequent years (Koenig 1992). While the original study took the cholera vaccine to be the treatment and tetanus-diphtheria toxoid to be the control we take children of the women who received the tetanus-diphtheria toxoid to be our treatment group, with the children of the mothers who received the cholera vaccine as the control group. We assume that the cholera vaccine administered to the mothers had no effect on their

children (any protective effect will reduce our estimates of the effect of the tetanus vaccination). We follow up with 12,048 of the children born over the period 1975-1979 to the women who participated in the trial using regular demographic surveillance data and a 1996 socioeconomic survey that covers all children still living in the area.

The analysis is complicated by the fact that, while the allocation of woman to tetanus-diphtheria toxoid or cholera vaccine was random, though conditional on agreeing to participate in the study, getting a second dose may suffer from selection bias. We adopt an intention to treat framework where we estimate the impact of being in the tetanus-diphtheria toxoid vaccine group relative to the cholera vaccine group independently of the number of doses received.

We also investigated the effect of treatment (the full two doses of the tetanus toxoid) on the treated. Estimating the effect of treatment on the treated means a loss of the randomized design and a need to include additional assumptions about why some women obtained a second dose, while some did not, to identify the effect. Under the assumption that adherence to the two dose treatment regime, versus receiving only one, dose was random, we found no statistical difference between mortality outcomes, or schooling attained, between one dose and two doses of the tetanus toxoid. Given this lack of difference, and the problems of identifying the effect of treatment on the treated, we focus on the intention to treat approach.

Some mothers are likely to have acquired immunity to tetanus from health providers prior to the intervention (there is no natural acquired immunity), though these efforts are not documented. A small baseline immunological study through finger prick blood sampling indicated some cases of immunity to tetanus in adults (Feeley, Curlin et al. 1979). From March 1978 tetanus toxoid was offered to pregnant women in part of the study area through this



program had very limited take up before 1980 (Fauveau 1994). Following our intention to treat framework we keep these women, and their children, in the sample for analysis.

The randomization of the treatment in the first dose, and our intention to treat framework based on this initial randomization, means that our treatment should be uncorrelated with any potential confounders. While we analyze the effect of immunization against tetanus in children, the unit of randomization was women, the mothers of the children, and not the children themselves. We control for this in our analysis by allowing for correlated outcomes for siblings.

Table 1 summarizes the data on the treatment and control groups. Between 1975 and 1979, 12,455 children were born in the Matlab surveillance area to 8,656 women who took part in the trial. About three quarters of the mothers received two doses of their treatment while about one quarter received only one dose. We take the treated group to be the 6,153 children whose mothers received either one or two doses of diphtheria-tetanus toxoid. We take the control group to be the 6,302 children of mothers who received the unsuccessful cholera vaccine.

We have data on the socioeconomic status of the women in our experiment from the 1982 Matlab Socioeconomic Census. We tested for a difference in characteristics between the treatment and control groups in terms of mothers' age at the time of the intervention, fathers' age at the time of the intervention, mothers' level of schooling, fathers' level of schooling, the families' land ownership, the number of living siblings at the time of the child's birth, the mortality rate among prior siblings, and the days between the intervention and the birth. The average of these variables in the treatment and control groups is shown in Table 2. As expected, none of the differences in these characteristics between the treatment and control groups is statistically significant at the 5% level.

We examine the school attainment of the children in our sample measured in the 1996 Matlab Socioeconomic Census. Table 3 shows the status of our children by the time of the 1996 Census. Just over half the children born between 1975 and 1979 are present in this survey. A little under a quarter of the children had died by 1996 while just over a quarter had migrated out of the area. In our analysis we focus on the status of surviving children. However, outmigration may lead to a sample selection bias in our results if it depends on the treatment of schooling outcomes, and in our analysis we correct for this selective migration.

We take education to be schooling in government, private or Non-Governmental Organization schools (most children attend government schools). We exclude education in the Madrasah system which focuses on religious instruction. Figure 1 shows the distribution of education for the sample of children present in the 1996 survey. A significant proportion of children in our sample have zero years of education. There are also peaks in educational attainment at five years, which matches completion of primary school, and at nine years which is close to completion of junior high (8 years) or high school (10 years) in Bangladesh.

## **Results**

We now examine the differential educational outcomes in the treatment and control groups. In table 4 we report ordinary least squares (OLS) regressions explaining the years of schooling attained by each child. Standard errors are Huber-White robust and clustered on mothers. This clustering allows for correlated outcomes between children born to the same mother, and controls for the fact that randomization was on mothers and not children. We find no significant effect of treatment on years of schooling in the full sample, as reported in column 1. In column 2 of table 4 we add covariates from the 1982 Socioeconomic Census, again finding no significant

effect for treatment. Adding covariates may lead to an efficiency gain by explaining some of the variation in outcomes, though because covariates are measured in 1982, and the intervention was in 1974, there is scope for an effect of treatment on these family characteristics, in which case the simple unadjusted estimates in column 1 may be preferable.

We now investigate if the effect differs for families of low socio economic status, which we take to be parents with no education (about half of our sample). In columns 3 and 4 of Table 4 we repeat the analysis for children whose parents have no education. We do find a significant effect of treatment at the 5% level, reported in column 3, with treated children receiving about a quarter of a year of additional education. This effect remains in the results reported in column 4 of Table 4, when we control for covariates, though it is significant only at the 10% significant level in this regression.

While the OLS results are suggestive of an effect, the distribution of education attainment plotted in Figure 1 shows that a significant proportion of children achieve no schooling. This suggests that OLS, and the assumption of a normal error term, may be inappropriate since educational attainment cannot be negative and many children are at this boundary. In Figure 2 we plot the difference in education attainment between treatment and control groups for the children of parents with no education. For each level of schooling the graph shows the difference between the percentage of the treated group attaining that level and the percentage of the control group attaining that level. Thus we find that children of parents with no education in the treatment group are about 3 percentage points less likely to have no schooling. However, this group of treated children is about 3 percentage points more likely to have achieved the junior high level of schooling (8 years of schooling).

To model the process shown in Figures 1 and 2 we use an ordered probit analysis with three discrete outcomes, zero years of schooling, 1-7 years of schooling, and 8 or more years of schooling. Each ordered probit contains a constant. Standard errors are Huber-White robust and clustered on mothers. As in the case of OLS, we report unadjusted estimates and adjusted estimates where we control for potential confounding variables. Table 5 reports results from this ordered probit. Again for the full sample treatment is not a significant predictor of schooling achievement. For the sample of children born to parents with no education, however, we do find a positive and significant coefficient on treatment.

This positive coefficient in the ordered probit model means that treatment tends to increase educational attainment for children from low socioeconomic backgrounds. To see the pattern of these gains we use the results from Table 5 to calculate the effect of treatment on the probability of being in each group. The results are reported in Table 6. Row 1 of Table 6 corresponds to column 3 of Table 5. The first row gives the effect on the probability of each outcome of maternal tetanus toxoid in an unadjusted ordered probit. We find that without additional covariates treatment reduces the probability of no schooling by about -0.047 with a corresponding increase in the probability of 1-7 years of schooling of about 0.018 and an increase in the probability of 8 or more years of schooling by 0.029. All these estimated effects are statistically significant at the 1% level. These effects cumulate to an increase in about 0.25 years of schooling in the treatment group relative to the control group. The second row of results in Table 6 corresponds to the effect of treatment in column 4 of table 9, where we correct for covariates. The rows below this give the marginal effects of the covariates on each outcome. As expected we find the estimated effect of maternal vaccination against neonatal tetanus is robust

to adding these additional control variables. The additional rows in Table 6 give the marginal effects of the covariates on schooling outcomes.

#### **4. Correcting the Schooling Outcomes for Selective Attrition Due to Out-Migration**

Our sample includes only children present in Matlab and covered by the 1996 Socioeconomic Census. Table 3 shows the status of our sample of children in this census. By 1996 about 20% of the original 12,455 children born have died. However, approximately another 26% have migrated out of the surveillance area and are lost to follow-up. This leaves about 54% of the original children in the sample for analysis (we have education data for a small number of children who are listed as having migrated out, which may reflect data mistakes, or children who have migrated but were temporarily visiting the site during the census).

We do not correct for mortality selection. We want to find the effect on education conditional on survival – this is the relevant effect for policy purposes. However, the high level of out-migration does pose a problem for estimation. While the educational attainment of the out-migrants is not observed it is an interesting outcome of the intervention; in principle we want to find the effect of the treatment on all children who survive, independently of whether they migrate out of the area or not. If migration is random it will not affect our results. It is more likely that treatment, and the health and educational benefits it provides, affects migration and that this produces a selection bias in our results based only on those who remain in survey area in 1996.

To address this problem we use a weighting approach to correct for selective outmigration (Fitzgerald, Gottschalk et al. 1998). Let  $A$  denote attrition from the sample, taking the value 1 if the child is not observed at the time of the survey and the value 0 if they are present

and observed. Let  $f(y|x)$  be the population density of  $y$ , the educational attainment, conditional on exogenous factors  $x$ , which is the object of interest, and let  $g(y|x, A = 0)$  be the density conditional on being in the sample in 1996, which is what we observe. It can be shown using Bayes rule that (Fitzgerald, Gottschalk et al. 1998):

$$f(y|x) = w(x, y)g(y|x, A = 0) \quad (1)$$

Where

$$w(x, y) = \left[ \frac{P(A = 0|x, y)}{P(A = 0|x)} \right]^{-1} \quad (2)$$

We can estimate the true population effect of the exogenous variables  $x$  on  $y$  by weighting the effect in the observed sample (where  $A = 0$ ), where the weights  $w(x, y)$  depend on the probability of being observed in the sample conditional on  $x$  relative to conditioning on both  $x$  and  $y$ . If migration depends on exogenous factors,  $x$ , but, given these, is independent of the schooling level,  $y$ , we would have

$$P(A = 0|x, y) = P(A = 0|x) \quad (3)$$

and the weights are the same on each observation in the sample. In this case we could simply analyze our data on children who remain in the sample, since the out-migration does not produce a bias. If, however, migration is selective and depends on the endogenous variable, schooling, this selectivity will bias our results if we fail to correct for it.

This weighting approach has been used in a panel setting in which attrition in each period may be a function of the endogenous variable in the previous period (Fitzgerald, Gottschalk et al.

1998). A simple test of endogenous selection in migration in our setting is to see if schooling affects attrition in the subsequent period. We cannot do this for migration before 1996, but we can look at migration of our observed sample after 1996. Table 7 reports the results of a probit model of outmigration after the Socioeconomic Census in mid 1996 but before the end of 1998, a period of about two and a half years. Given that we observe schooling for this sample we can include it in the regression. We find that children who are more highly schooled in 1996 are more likely to migrate out over the subsequent two and half years. Each year of schooling increases the probability of out-migration over this period by about 0.03. This suggests that outmigration is correlated with schooling, even conditioning on the exogenous variables, and that the unweighted estimators of the effect of treatment on education, reported in tables 4, 5 and 6, are biased.

We wish to correct for the selection bias. To do this we need to estimate the probability of being in the sample, conditional on schooling. In order to identify this effect we have to make some assumptions about the nature of the migration process. We assume migration follows a Cox proportional Hazard model (Cox 1972; Lee and Wang 2003), with a time varying baseline migration hazard and proportional hazards that depend on treatment and school attainment but are time invariant. We estimate this proportional hazard model for the two year period 1/1/1997 to 1/1/1999 using schooling data from the 1996 Socioeconomic Survey. The estimation of the hazard uses daily data where a child is at risk each day if they have not yet migrated out or died. The results are reported in the top portion of table 8. This gives us an estimate of the relative hazard by child characteristics. Our estimates suggest that the treated are somewhat less likely to migrate out, though the effect is not statistically significant. The migration hazard does rise with years of schooling and this effect is statistically significant in the sample of children whose

parents had no schooling. The probability of remaining in the sample, by year of birth, for baseline children, who are in the control group (mothers did not receive tetanus toxoid) and have zero years of schooling, are reported in the bottom portion of table 8.

Under the assumption that the relative hazard rates do not vary with age we can apply these results on the baseline probability of migration, and proportional hazards to estimate the probability of out-migration before 1996 by the child's characteristics. Details of this calculation are given in an appendix. Using these probabilities we can then calculate the weights required to correct for selective outmigration using equation (2).

The weights we calculate vary from a low of 0.92 to a high of 1.12 in the full sample. The low weights are associated with young children who have low levels of schooling; they are unlikely to have migrated out. The high weights are associated with older children with high levels of schooling who are more likely to migrate, and who are therefore underrepresented in our unweighted regression analysis in the previous section.

Table 9 reports the results for regressions in which we weight observations, using the formula in equation (2) to correct for selective migration. We give the coefficient for the effect of treatment on the probability of achieving each level of schooling. These results correspond to the unweighted regressions without additional controls reported in row 1 of Table 6. Weighting has little effect on the estimates; we find similar magnitudes to those reported in Table 6 and the results remain significant at the 5% level. We do not report weighted estimates for regressions with additional control variables. Some of these additional variables are missing in the data for some observations, creating an additional potential selection problem.



The model set out in this section controls for migration on observables, and on schooling attained. We assume that given these migration is random and our results would be undermined by migration selection on unobserved characteristics. The results in this section also rely on the assumption that the relative hazard of migrating is constant over time. It may be that while schooling (or the latent variable that is associated with it) is an important determinant of migration for older children, it is less important for younger children. In this case our weighting scheme will be too extreme and the unweighted estimates may be closer to the actual effects. However, the fact that our estimates change very little, even with this extreme weighting suggests that our results are not very sensitive to selection.

Our results on mortality and schooling are consistent with the idea that without maternal vaccination against tetanus about 7% of children develop neonatal tetanus and about half of these children die while the other half survive, but have long term problems in physical or cognitive development that lead to low educational achievement.

### **Cost Benefit Analysis**

The recommended vaccination regime is to give pregnant woman three doses of tetanus toxoid. The full cost per woman of tetanus immunization using this three dose regime has been estimated at \$1.19 in Pakistan in a community study (Griffiths, Wolfson et al. 2004), and between \$3.28 and \$4.06 for scaling up of existing efforts in South Asia and Sub-Saharan Africa (Brenzel, Wolfson et al. 2006). Based on our ordered probit estimates, the Matlab intervention increases schooling attainment by about 0.25 years on average (though the effect was concentrated in a small number of children) for children from low socio-economic backgrounds. Three doses of the toxoid to the mother provide effective immunization to children born in at

least the next five years, and likely much longer. Total fertility rates in all countries exceed one and in some African countries exceed five. We assume at least one birth per woman during her immunized period, which is probably conservative. If a woman has one protected child this implies a cost of between \$4.76 and \$16.16 per additional year of schooling, though the cost per child in high fertility settings is likely much lower. This compares with estimated costs per year of schooling increased of \$3.50 for deworming, \$36 for a school feeding program, and \$99 for free uniforms (Kremer 2003). This suggests that maternal tetanus vaccination can be considered as a cost effective educational intervention for children from low socio-economic status households.

Current efforts to eliminate neonatal tetanus focus on community based interventions in high risk settings. These settings usually lack access to routine health care and are likely predominately comprised of low socio-economic status households. As well as seeing reductions in mortality due to these interventions, our results from Matlab suggest we should also see improved schooling outcomes.

## **Discussion**

We find significant effects of maternal tetanus vaccination on the schooling outcomes of children. A strength of our study is the randomized nature of the intervention. While randomization was on the mother rather than the child, we have a large number of randomized groups - those children born to the same mother. The large number of groups makes the results more robust than randomized group level interventions with a small number of groups, such as villages, where there may be significant variation in covariates between the treatment and control populations. We also use appropriate and robust statistical techniques, allowing for the discrete nature of the educational outcome, the correlation between outcomes between children born to the same mother, and the possibility of selective migration.

In terms of the internal validity, the causal effect of the intervention on the population in Matlab, we have two potential problems. The first is that we see effects only on those children whose mothers agreed to participate in the study, not the whole population, though randomization of the intervention should produce valid estimates for the effect on this group. A second issue is that best practice is for the outcome variable in a study to be specified prior to the randomization taking place. This was not the case here; the original purpose of the 1974 trial was to look at effects of the cholera vaccine, not the tetanus toxoid, and schooling was not thought of as an endpoint. We also find effects only on a subgroup that was specified *ex post* rather than *ex ante*. Randomization of a tetanus toxoid intervention, to study its educational effects, would not be ethically justified, since we now know the intervention to have substantial mortality benefits. It therefore appears we have to live with these issues and make the best we can of the available data.

There is also an issue of external validity. What would we expect to happen if maternal tetanus immunization was extended to other areas? The incidence of neonatal tetanus in the Matlab site, while high, was not exceptional. Neonatal tetanus mortality rates of between 23 and 82 per 1000 births have been observed in un-immunized populations (very high rates of neonatal tetanus are observed particularly when unhygienic methods are usually employed, such as applying cow dung to the umbilical stump), and large reductions in neonatal mortality have been observed worldwide as maternal tetanus immunization has been rolled out (Meegan, Conroy et al. 2001; Roper, Vandelaer et al. 2007). Neonatal tetanus is now rare in developed countries but it still generates a high disease burden in Africa and Asia (World Health Organization 2006). Globally as of 2008, 73% of pregnant woman received tetanus toxoid vaccination, up from 63% in 2002 (World Health Organization 2009). As of 2002 there were an estimated 30,000 maternal and 180,000 neonatal deaths from tetanus each year that could have been avoided with complete maternal vaccine coverage (WHO and UNICEF 2005).

The magnitude of the schooling effect depends on the effects on neonatal tetanus on children's physical and cognitive development, and the effect of this development on their school attendance and progression. Our results suggest large effects in families with low socioeconomic status where poor nutrition may magnify the effect of infection. The results therefore depend on the status of the families in which the intervention is implemented. In our study based on births in the 1970s about half the children had parents with no schooling. Increasing enrollment rates means that the education level of younger cohorts tends to be higher than older cohorts, but there are still significant numbers of parents in the world who lack schooling. For example, as of 2000, in India 48% of women, and 25% of men in the age group 15-49 had no schooling (Lutz, Goujon et al. 2007). In addition, the schooling achieved depends on the educational resources

available. In settings that have different schooling systems from those in Matlab during the 1980s and 1990s we may see different effects.

While we see significant effects on schooling outcomes for children from whose parents had no schooling in the treatment group, and there is substantial evidence of an economic return to schooling in the form of higher wages (Bloom, Canning et al. 2005), we would like to see direct evidence of increases in incomes when these more children with more schooling enter employment in order to infer an economic impact of the intervention. However, at present we lack follow up data on incomes.

Our results make a case for investments in maternal tetanus vaccination as a method of improving schooling, and eventually economic, outcomes. There is already a strong case for complete vaccination coverage and elimination of maternal and neonatal tetanus as a cost effective health intervention (WHO and UNICEF 2005). Despite that, coverage is still far from complete and the evidence provided here may strengthen the case for increased funding. Similar economic arguments are possible for other vaccines (Bärnighausen, Bloom et al. 2009).

**Table 1**  
**The Treatment and Control Groups**

		Mothers	Children born 1975-1979
Treatment and Control Groups 1974	Control (1-2 doses cholera vaccine)	4,376	6,302
	Treated (1-2 doses tetanus toxoid)	4,280	6,153
	Total	8,656	12,455
Status of Children 1996	Dead		2,428
	Missing from survey		3,215
	Schooling data available		6,812
	Total	8,656	12,455

**Table 2**  
**Average Maternal Characteristics by Group**

	Treated	Control	t-test of difference in means
Mother's age at 7.1.74 (years)	24.3	24.3	0.35
Father's age at 7.1.74 (years)	34.4	34.5	0.55
Mother's years of Schooling	1.10	1.01	1.84
Father's years of Schooling	2.55	2.44	1.37
Land Owned 1982	0.856	0.846	0.35
Living Children at first birth after 1974	2.35	2.37	0.45
Child Mortality Ratio at first birth after 1974	0.254	0.267	1.83
Days to first Birth	762.8	762.5	0.14

**Table 3**  
**Status of Children in the 1996 Socioeconomic Census**

Status in 1996 survey	Number	Percentage
Dead	2428	19.5
Migrated	3223	25.9
Present	6804	54.6
Schooling data available	6812	54.8
Total	12,455	100



**Table 4**  
**The Effect of Maternal Tetanus Vaccination on Children's Years of Schooling**

Dependent Variable: Years of Schooling  
 Estimation Method: Ordinary Least Squares

	Full sample		Parents with no schooling	
Constant	5.244** (73.2)	3.91** (12.15)	3.372** (34.8)	3.433** (7.09)
Treated	0.072 (0.72)	0.024 (0.26)	0.278* (1.98)	0.267 (1.91)
Female		-0.146 (1.72)		-0.272* (2.14)
Mother's Schooling (years)		0.418** (15.6)		
Fathers Schooling (years)		0.328** (19.6)		
Mother's Age at Birth (years)		-0.006 (0.51)		-0.006 (0.34)
Land Owned 1982 (acres)		0.073* (2.09)		0.059 (0.97)
Living Children at first birth in period		0.177** (4.34)		0.182** (2.91)
Child Mortality Ratio pre-first birth in period		-0.080 (0.40)		-0.129 (0.43)
Born 1976		-0.125 (0.75)		0.008 (0.03)
Born 1977		-0.340* (2.13)		-0.150 (0.61)
Born 1978		-0.210 (1.31)		0.138 (0.55)
Born 1979		-0.617** (3.96)		0.021 (0.09)
R <sup>2</sup>	0.0001	0.241	0.0018	0.015
N	6812	5844	2663	2663

Standard errors are Huber-White robust and clustered on mothers, t-statistics in parenthesis.  
 We indicate significance at the 5 percent (\*) and 1 percent levels (\*\*).

**Table 5**  
**The Effect of Maternal Tetanus Vaccination on Children's Schooling Attainment**

Dependent Variable: Years of schooling in three categories.  
 Estimation Method: Ordered Probit Model

	Full sample		Parents with no schooling	
Treated	0.035 (1.17)	0.038 (1.15)	0.123*** (2.60)	0.119** (2.51)
Female		-0.108*** (3.51)		-0.163*** (3.69)
Mother's Schooling (years)		0.145*** (13.5)		
Fathers Schooling (years)		0.113*** (17.9)		
Mother's Age at Birth (years)		0.000 (0.03)		0.002 (0.33)
Land Owned 1982 (acres)		0.032** (2.41)		0.029 (1.33)
Living Children at first birth in period		0.058** (4.03)		0.054** (2.55)
Child Mortality Ratio pre-first birth in period		-0.084 (1.22)		-0.097 (0.95)
Born 1976		0.017 (0.29)		-0.016 (0.18)
Born 1977		-0.028 (0.50)		-0.058 (0.68)
Born 1978		0.008 (0.13)		0.020 (0.23)
Born 1979		-0.144 (2.66)		0.001 (0.01)
Pseudo R <sup>2</sup>	0.0001	0.117	0.001	0.011
N	6812	5844	2663	2663

Coefficients are parameters in ordered probit model.  
 Standard errors are Huber-White robust and clustered on mothers, t-statistics in parenthesis.  
 We indicate significance at the 5 percent (\*) and 1 percent levels (\*\*).

**Table 6**  
**Marginal Effect of Maternal Tetanus Vaccination on**  
**Children's Probability of School Attainment**

**Parents with no schooling**

	School Attainment		
	No Schooling	1-7 Years of School	8 or More Years of School
Treated (no controls)	-0.047** (2.60)	0.018** (2.59)	0.029** (2.58)
Treated (with controls)	-0.045** (2.51)	0.017** (2.50)	0.028** (2.49)
Female	0.062*** (3.69)	-0.024*** (3.61)	-0.038*** (3.65)
Mother's Age at Birth (years)	-0.001 (0.33)	0.000 (0.33)	0.000 (0.33)
Land Owned 1982 (acres)	-0.011 (1.33)	0.004 (1.33)	0.007 (1.33)
Living Children at first birth in period	-0.020** (2.56)	0.008** (2.53)	0.013** (2.54)
Child Mortality Ratio pre-first birth in period	0.037 (0.95)	-0.014 (0.95)	-0.023 (0.95)
Born 1976	0.006 (0.18)	-0.002 (0.18)	-0.004 (0.18)
Born 1977	0.022 (0.68)	-0.009 (0.66)	-0.013 (0.70)
Born 1978	-0.008 (0.23)	0.003 (0.24)	0.005 (0.23)
Born 1979	-0.0003 (0.01)	0.0001 (0.01)	0.0002 (0.01)

Coefficients are marginal effects on probabilities from the ordered probit model. Standard errors are Huber-White robust and clustered on mothers, t-statistics in parenthesis. We indicate significance at the 5 percent (\*) and 1 percent levels (\*\*).

**Table 7**  
**Children's Migration after after 1996 Socioeconomic Census**

Dependent Variable: Outmigration after 1996 Census but before January 1, 1999:  
 Estimation Method: Probit Model

	Full sample		Parents with no schooling	
Treated	-0.009 (1.21)	-0.010 (1.17)	-0.006 (0.46)	-0.007 (0.62)
Child's Years of Schooling in 1996	0.031** (3.38)	0.030** (3.04)	0.033** (2.63)	0.034** (2.75)
Female		0.050** (6.24)		0.052** (4.41)
Mother's schooling (years)		0.002 (0.96)		
Fathers schooling (years)		-0.002 (1.57)		
Mother's Age at Birth (years)		-0.001 (0.68)		0.0001 (0.09)
Land Owned 1982 (acres)		-0.002 (0.50)		0.004 (0.78)
Living Children at first birth in period		0.0002 (0.05)		0.001 (0.11)
Child Mortality Ratio pre-first birth in period		-0.013 (0.75)		-0.016 (0.66)
Born 1976		0.005 (0.38)		0.021 (0.89)
Born 1977		-0.020 (1.41)		-0.002 (0.09)
Born 1978		-0.007 (0.46)		0.008 (0.34)
Born 1979		-0.005 (0.35)		0.020 (0.87)
Pseudo R <sup>2</sup>	0.003	0.014	0.004	0.017
N	6789	5825	2652	2652

Exclude those who died before the 1996 census.

Coefficients are marginal effects on the probability of migrating

Standard errors are Huber-White robust and clustered on mothers, t-statistics in parenthesis.

We indicate significance at the 5 percent (\*) and 1 percent levels (\*\*).

**Table 8**  
**The Effect of Schooling on Outmigration**

Estimation Method: Cox Proportional Hazard Model

		Full sample	Parents with no schooling
Relative hazard rate Estimated on data for 1997-1998	Treated	0.906 (1.24)	0.946 (0.43)
	Years of Schooling	1.020 (1.93)	1.038* (2.04)
	N	6696	2613
	Outmigrants	675	248
Baseline probability of remaining in the 1996 sample by birth year	1975	0.611	0.632
	1976	0.664	0.691
	1977	0.693	0.725
	1978	0.724	0.772
	1979	0.778	0.808

Breslow method for ties.

Standard errors are Huber-White robust and clustered on mothers, t-statistics in parenthesis.

Base line is control group with zero years of education.

We indicate significance at the 5 percent (\*) and 1 percent levels (\*\*).

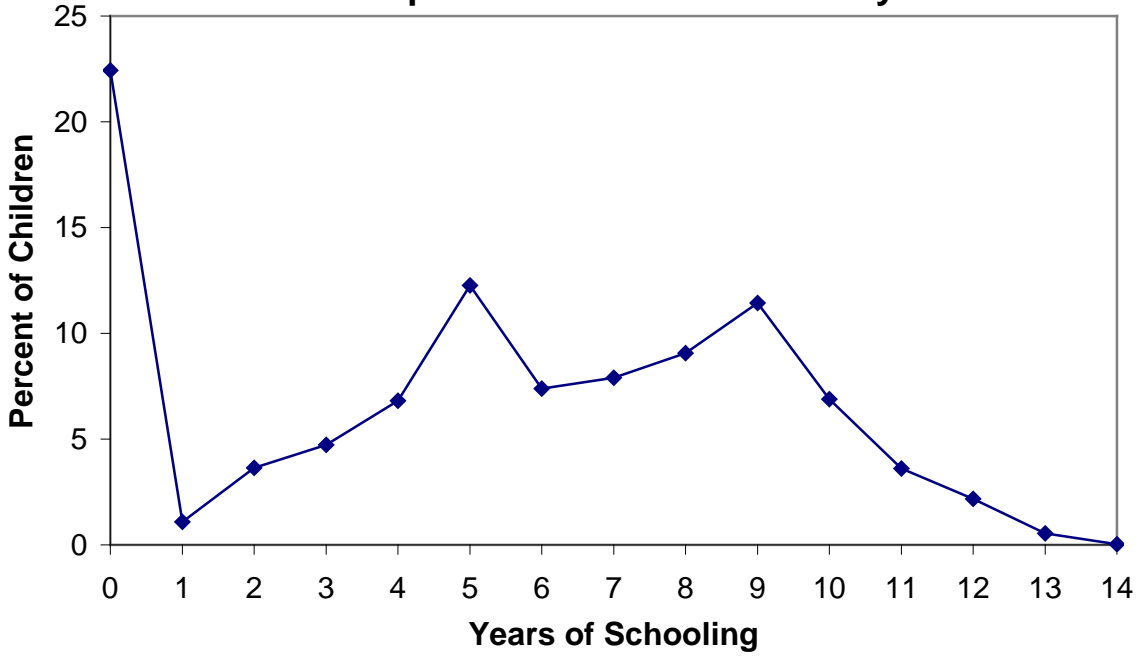
**Table 9**  
**Marginal Effect of Maternal Tetanus Toxoid on Children's School Attainment**  
**Corrected for Selective Outmigration**  
**Parents with no schooling**

Estimation Method: Weighed Ordred Probit Model

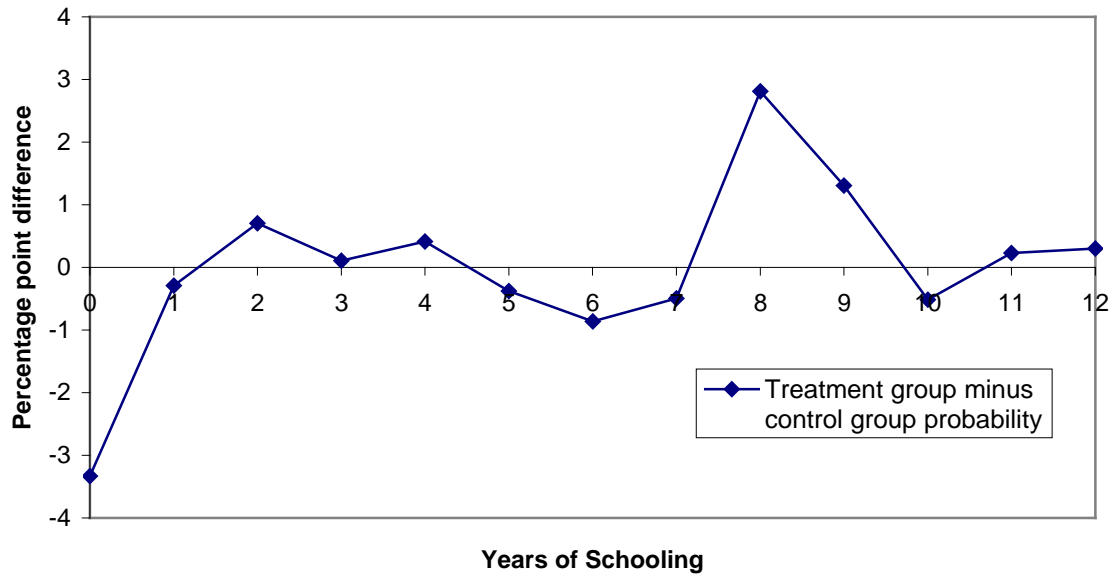
	Probability of No schooling	Probability of 1-7 Years of school	Probability of 8 or more years of school
Treated	-0.045** (2.55)	0.015** (2.53)	0.030** (2.53)

Estimates are treatment effects on probabilities from an ordered probit model. Controls are child's sex, mother's age at birth, parents' land holding, living prior children, dead prior children ratio, birth year dummies. Weighting adjusts for sample attrition. 2663 observations.

**Figure 1**  
**Distribution of School Attainment**  
**of Sample Children in 1996 Survey**



**Figure 2**  
**Differential Schooling Attained by Treatment**  
**Parents with No schooling**





## References

- Bärnighausen, T., D. E. Bloom, et al. (2009). The Economic Case for Expanding Vaccination Coverage of Children. Program on the Global Demography of Aging, Working Paper 45, Harvard Center for Population and Development Studies
- Behrman, J. R., M. C. Calderon, et al. (2009). "Nutritional Supplementation of Girls Influences the Growth of their Children: Prospective Study in Guatemala." American Journal of Clinical Nutrition **90**: 1372-1379.
- Black, R. E., D. H. Huber, et al. (1980). "Reduction of neonatal tetanus by mass immunization of non-pregnant women: duration of protection provided by one or two doses of aluminium adsorbed tetanus toxoid." Bulletin of the world Health Organization **58**: 927-930.
- Black, R. E., S. S. Morris, et al. (2003). "Where and why are 10 million children dying every year?" Lancet **361**(9376): 2226-2234.
- Bleakley, H. (2003). "Disease and Development: Evidence from the American South." Journal of the European Economic Association **1**(2-3): 376-386.
- Bleakley, H. (2006). Malaria In The Americas: A Retrospective Analysis Of Childhood Exposure, UNIVERSIDAD DE LOS ANDES-CEDE.
- Bleakley, H. (2010). "Malaria Eradication in the Americas: A Retrospective Analysis of Childhood Exposure." American Economic Journal: Applied Economics **2**(2): 1-45.
- Bloom, D. E. and D. Canning (2000). "The health and wealth of nations." Science **287**: 1207-1208.
- Bloom, D. E., D. Canning, et al. (2005). "The Value of Vaccination." World Economics **6**(3): 15-39.
- Breiman (2004). "Effect of infant immunisation on childhood mortality in rural Bangladesh: analysis of health and demographic surveillance data." Lancet **364**(9452): 2204-2211.
- Brenzel, L., L. J. Wolfson, et al. (2006). Vaccine-Preventable Diseases. Disease Control Priorities in Developing Countries. D. T. Jamison, J. G. Breman, A. R. Measham et al. New York, Oxford University Press: 389-412.
- Cox, D. R. (1972). "Regression Models and Life-Tables." Journal of the Royal Statistical Society. Series B. **34**(2): 187-220.
- Curlin, G. T., R. J. Levine, et al. (1978). "Immunological aspects of a cholera toxoid field trial in Bangladesh." Dacca, Cholera Research Laboratory, Scientific Report No. 8.

- Cutler, D., W. Fung, et al. (2010). "Early-Life Malaria Exposure and Adult Outcomes: Evidence from Malaria Eradication in India." American Economic Journal: Applied Economics **2**(2): 72-94.
- Fauveau, V. (1994). Neonatal and Perinatal Health in Matlab. Matlab: Women, Children and Health. V. Fauveau. Dhaka, The International Centre for Diarrhoeal Disease Research, Bangladesh.
- Feeley, J. C., G. T. Curlin, et al. (1979). "Response of children in Bangladesh to adult-type tetanus-diphtheria toxoid (Td) administered during a field trial of cholera toxoid." Journal of Biological Standardization **7**(3): 249-252.
- Fitzgerald, J., P. Gottschalk, et al. (1998). "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics." Journal of Human Resources **33**(2): 251-299
- Griffiths, U. K., L. J. Wolfson, et al. (2004). "Incremental cost-effectiveness of supplementary immunization activities to prevent neonatal tetanus in Pakistan." Bull World Health Organ **82**(9): 643-651.
- Hoddinott, J., J. Maluccio, et al. (2008). "Effect of a nutrition intervention during early childhood on economic productivity in Guatemalan adults." Lancet **371**(9610): 411-416.
- Jukes, M. C. H., M. Pinder, et al. (2006) Long-Term Impact of Malaria Chemoprophylaxis on Cognitive Abilities and Educational Attainment: Follow-Up of a Controlled Trial. PLOS Clin Trial **1**(4): **Volume**, DOI: e19. doi:10.1371/journal.pctr.0010019
- Koenig, M. A. (1992). "Mortality Reductions from Measles and Tetanus Immunization: A Review of the Evidence." World Bank Policy Research Working Paper WPS868.
- Koenig, M. A., D. Bishai, et al. (2001). "Health interventions and health equity: the example of measles vaccination in Bangladesh." Population and Development Review **27**(2): 283-302.
- Koenig, M. A., N. C. Roy, et al. (1998). "Duration of Protective Immunity Conferred by Maternal Tetanus Toxoid Immunization: Further Evidence from Matlab, Bangladesh." American Journal of Public Health **88**(6): 903-907.
- Kremer, M. (2003). "Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons." American Economic Review **93**(2): 102-106.
- Lee, E. T. and J. W. Wang (2003). Statistical methods for survival data analysis. New York, John Wiley and Sons.
- Lucas, A. M. (2010). "Malaria Eradication and Educational Attainment: Evidence from Paraguay and Sri Lanka." American Economic Journal: Applied Economics **2**(2): 46-71.

Lutz, W., A. Goujon, et al. (2007). "Reconstruction of population by age, sex and level of educational attainment of 120 countries for 1970-2000." Vienna Yearbook of Population Research **2007**: 193-235.

Maluccio, J. A., J. Hoddinott, et al. (2009). "The Impact of Nutrition During Early Childhood on Education among Guatemalan Adults." Economic Journal **119**: 734-763.

Meegan, M. E., R. M. Conroy, et al. (2001). "Effect on neonatal tetanus mortality after a culturally-based health promotion programme." Lancet **358**(9282): 640-41.

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

Roper, M. H., J. H. Vandelaer, et al. (2007). "Maternal and neonatal tetanus." Lancet **470**(9603): 947-1959.

Subramanian, S. V., L. K. Ackerson, et al. (2009). "Association of Maternal Height With Child Mortality, Anthropometric Failure, and Anemia in India." Journal of the American Medical Association **301**(16): 1691-1701.

WHO and UNICEF (2005). Investment Case: Neonatal Tetanus Elimination. Submission to GAVI., [http://www.gavialliance.org/resources/16brd\\_06\\_MNT\\_investment\\_case.pdf](http://www.gavialliance.org/resources/16brd_06_MNT_investment_case.pdf).

World Health Organization (2006). "Tetanus Vaccine: WHO position paper." Weekly Epidemiological Record **81**: 198-208.

World Health Organization (2009). WHO vaccine-preventable diseases: monitoring system. 2009 global summary. Geneva, World Health Organization, Department of Immunization, Vaccines and Biologicals.

# The Effect of Maternal Tetanus Immunization on Children's Educational Attainment in Matlab, Bangladesh

## Appendix: Correctiong for out migration

There is a potential mortality selection effect, the children who survive to 1996 may be better, or worse, in terms of health and potential educational outcomes than those that die. We do not correct for this mortality selection. We want to find the effect on education conditional on survival.

However, the high level of out-migration does pose a problem for estimation. While the educational attainment of the out-migrants is not observed it is an interesting outcome of the intervention; in principle we want to find the effect of the treatment on all children who survive, independently of whether they migrate out of the area or not. The key issue is if there is selection which is correlated with our outcome variable, education, which we do not observe for the children who migrate. To address this problem we adapt the approach outlined by Fitzgerald, Gottschalk et al. (1). Let  $A$  denote attrition from the sample, taking the value 1 if the child is not observed at the time of the survey and the value 0 if they are present and observed. Let  $f(y|x)$  be the population density of  $y$  conditional on exogenous factors  $x$ , which is the object of interest, and let  $g(y|x, A = 0)$  be the density conditional on being in the sample in 1996, when we observe the outcome. Then

$$g(y|x, A = 0) = \frac{g(y, A = 0|x)}{P(A = 0|x)} \quad (4)$$

$$= \frac{P(A = 0|x, y) f(y|x)}{P(A = 0|x)} \quad (5)$$

$$= \frac{f(y|x)}{w(x, y)} \quad (6)$$

Where

$$w(x, y) = \left[ \frac{P(A = 0|x, y)}{P(A = 0|x)} \right]^{-1} \quad (7)$$

Hence

$$f(y|x) = w(x, y)g(y|x, A = 0) \quad (8)$$

We can estimate the true population effect of the exogenous variables  $x$  on  $y$  by weighting the effect in the observed sample, where the weights  $w(x, y)$  depend on the probability of being observed in the sample conditional on  $x$  relative to conditioning on both  $x$  and  $y$ . If migration depends on exogenous factors,  $x$ , but given these is independent of the educational attainment,  $y$ , we have

$$P(A = 0|x, y) = P(A = 0|x) \quad (9)$$

and the weights are the same on each observation in the sample. In this case we can simply analyze our data on children who remain in the sample, since the out-migration does not produce a bias. If, however, migration is selective and depends on the endogenous variable, schooling, this selectivity may bias our results, if we fail to correct for it.

Fitzgerald, Gottschalk et al. (1) use this approach in a panel setting in which attrition in each period may be a function of the endogenous variable in the previous period. They show that a simple test of endogenous selection is to see if the endogenous variable affects attrition in the subsequent period. We cannot do this for migration before 1996, but we can look at migration of our observed sample after 1996. We assume that there is a latent variable that drives both education and migration, so that migration in the early years is correlated with attained schooling even though this schooling is not achieved until later in life.

Let  $S(t, x, y)$  denote the probability that a child with characteristics  $x$  and education  $y$  (or its latent variable equivalent if educational achievement has not yet been realized) is still living in the survey area at age  $t$ . We can define the hazard of migrating out of the area at age  $t$  as

$$h(t, x, y) = -\frac{dS(t)/dt}{S(t)} = -\frac{d}{dt} \log S(t) \quad (10)$$

We now assume a Cox proportional hazard model (2, 3):

$$h(t, x, y) = h_0(t)h_1(x, y) \quad (11)$$

so that the effects of the child's characteristics and age are separable. The hazard at age  $t$  is equal to a baseline hazard  $h_0(t)$  and a relative hazard  $h_1(x, y)$ . This implies that the cumulative hazard of migrating out by age  $T$  is

$$-\ln \mathbb{P}(A = 0 | T, x, y) = \int_{t=0}^T h(t, x, y) dt = h_1(x, y) \int_0^T h_0(t) dt = \sigma_0(T)h_1(x, y) \quad (12)$$

Where  $\sigma_0(T)$ , the cumulative baseline hazard of migration by age  $T$ , is a function that depends on age but is independent of the child's other characteristics.

Equation (9) cannot be estimated directly since we do not observe the education of the children who have migrated out. However we have data on migration after the 1996 census. We estimate the Cox proportional hazard model for the two year period 1/1/1997 to 1/1/1999 taking  $h_1(x, y) = \exp(\beta x + \gamma y)$  where  $x$  indicates treatment and  $y$  indicates education, in years. The estimation of the hazard uses daily data where a child is a risk each day if they have not yet migrated out or died. Out-migrants contribute to the estimation of the hazard while those who die are no longer at risk of migration on subsequent days. The results are reported in the top portion of Table 8 in the paper. This gives us an estimate of the relative hazard function  $\hat{h}_1(x, y)$ . Our estimates suggest that the treated are somewhat less likely to migrate out, though the effect is not statistically significant. The migration hazard does rise with years of schooling and this effect is statistically significant in the sample of children whose parents had no education.

Under the assumption that the relative hazard rate is age invariant we can apply these results to estimate the probability of out-migration before 1996. We can estimate the cumulative baseline hazard at age  $T$  by the condition

$$\sum_{i:a_i=T} P(A_i = 0 | T, x_i, y_i) / N_T = \sum_{i:a_i=T} \exp[-\sigma(T)h_1(x_i, y_i)] / N_T \quad (13)$$

Where we sum over children  $i$  who are age  $a_i = T$  in 1996. Condition (10) is just that the expected proportion of children still in the area at time  $T$  is the sum of the conditional probabilities of each child being present. We can get an estimate of the left hand side of equation (10) by replacing the average probability of being in the sample at age  $T$  by the observed fraction of children who remain in the sample at a age  $T$ . In addition note that

$$\begin{aligned} \exp[-\sigma(T)h_1(x_i, y_i)] &= \exp[-\sigma(T)\bar{h}_{1T}(x_i, y_i)]\exp[-\sigma(T)(h_1(x_i, y_i) - \bar{h}_{1T}(x_i, y_i))] \\ &\approx [1 - \sigma(T)(h_1(x_i, y_i) - \bar{h}_{1T}(x_i, y_i))]\exp[-\sigma(T)\bar{h}_{1T}(x_i, y_i)] \end{aligned} \quad (14)$$

Where we use the approximation  $e^z \approx 1 + z$  for  $z$  small. Replacing the expected proportion of children age  $T$  in 1996 in the left hand side of equation (10) with the observed proportion  $\bar{S}_T$  and using equation (11) to approximate the right hand side of equation (10) we can derive

$$\bar{S}_T = \sum_{i:a_i=T} \exp[-\hat{\sigma}(T)h_1(x_i, y_i)] / N_T \approx \exp[-\hat{\sigma}(T)\bar{h}_{1T}(x_i, y_i)] \quad (15)$$

This gives us the estimated cumulative baseline hazard  $\hat{\sigma}(T)$  defined by

$$\hat{\sigma}(T) = -\log(\bar{S}_T) / \bar{h}_{1T}(x_i, y_i) \quad (16)$$

Which implies

$$e^{-\hat{\sigma}(T)} = \bar{S}_T^{1/\bar{h}_{1T}(x_i, y_i)} \quad (17)$$

In practice we approximate the average of the relative hazards,  $\bar{h}_{1T}(x_i, y_i)$  aged  $T$  in the population by the average of those in the sample aged  $T$  in 1996.<sup>1</sup> These give the estimated probability that a baseline child (one in the control group who has zero years of schooling) has migrated out of the area by 1996. These figures for baseline are lower than the actual migration rates since children who with no education are less likely to migrate than an average child. For each child we can compute the probability of being in the sample in 1996 given their age  $T$ , and characteristics as:

$$\hat{P}(A_i = 0|T, x, y) = \exp[-\hat{\sigma}(T)\hat{h}_1(x, y)] = \bar{S}_T^{h_1(x, y)/\bar{h}_{1T}(x, y)} \quad (18)$$

For observations with high relative hazards of migrating, the exponent in equation (15) is greater than one, reducing the probability of remaining in the sample (since  $\bar{S}_T < 1$ ). The probability of remaining in the sample, by year of birth, for baseline children, who are in the control group and have zero education, are reported in the bottom portion of table 8.

We can estimate the probability of having migrated out at age  $T$  conditional on treatment,  $x_i$ , without any problem of observably. In this case we simply take the probability as the

---

<sup>1</sup> If we have only one age group, this approximation would not affect the weighted estimates. In this case every observation has the same age, and baseline cumulative hazard rate, in 1996, and so this part of the hazard function does not affect the weighted estimates, which depend only on relative weights. We prefer to use the approximate baseline cumulative hazard rate estimates so we can combine children at different ages in 1996 in the same weighted regression.

proportion of children in group  $x$  who have migrated out by age  $T$ ,  $\bar{S}(T, x)$ . We then have the weights for children of age,  $T$ , exogenous characteristics,  $x$ , and years of schooling,  $y$ :

$$w(T, x, y) = \left[ \frac{P(A = 0|T, x, y)}{P(A = 0|T, x)} \right]^{-1} = \left[ \frac{\exp[-\hat{\sigma}(T)\hat{h}_1(x, y)]}{\bar{S}(T, x)} \right]^{-1} \quad (19)$$

The weights vary from a low of 0.92 to a high of 1.12 in the full sample. The low weights are associated with children who have low education, they are unlikely to migrate out. The high weights are associated with children with high levels of education who are more likely to migrate, and who are therefore underrepresented in our regression analysis in the previous section.

The weighted results show in Table 9 in the paper use this weighting scheme. The results in this section rely on the assumption that the relative hazard of migrating is constant over time. It may be that while education (or the latent variable that is associated with it) is an important determinant of migration for older children, it is less important for younger children. In this case our weighting scheme will be too extreme and the unweighted estimates may be closer to the actual effects. However, the fact that our estimates change very little, even with this extreme weighting suggests that our results are not very sensitive to selection.

1. Fitzgerald J, Gottschalk P, Moffitt R. An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics. *Journal of Human Resources*. 1998;33(2):251-99
2. Cox DR. Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B*. 1972;34(2):187-220.
3. Lee ET, Wang JW. *Statistical methods for survival data analysis*. 3 ed. New York: John Wiley and Sons; 2003.