Online Appendix

Contents

List of Figures

List of Tables

A Data Appendix

A.1 Malawi Longitudinal Survey of Families and Health (MLSFH):

Our main analysis uses data from 4 waves of the MLSFH: 2004, 2006, 2008, and 2010.[2] We also use the 1998 and 2001 waves of the survey to impute savings. The MLSFH collected GPS coordinates of households in 2004, and the survey team also conducted HIV testing in 2004-2008 waves of the survey.[3] Figure **??** shows the distribution of spatial characteristics and location (within the circle) of respondents for each region of the survey.

- *Sample selection:* The sample selection process used the 2006 wave as the baseline year because many key variables, particularly savings, subjective expectations, and earnings, were collected only starting 2006. Since GPS coordinates were collected only in 2004, the starting sample for our analysis was individuals from the 2004 sample that were reinterviewed in 2006, which consisted of 2546 individuals (see Figure 2 from Kohler et al. (2015) for the MLSFH Sample Flow). From this sample, we were able to obtain accurate GPS data for 2095 individuals.[4]

  Our analytical sample consists of the set of respondents with GPS data that are present for all three years of the survey, from 2006-2010, which is 1379 individuals in 2006. This was done in order to ensure that estimates are not biased due to changing sample composition throughout the years. For outcomes with 2004 data, we use the available data for those respondents in the analytical sample that were interviewed in 2004, but still include the respondents that have missing data in 2004. This is done to maximize the sample size using the full set of outcomes of interest that are available (since savings, earnings, and subjective expectations were only added to the survey module beginning 2006). The results using a balanced panel starting with 2004 are show in Appendix Tables A2 and A3. Finally, for imputing savings before 2006, we incorporate the 1998 and 2001 waves of the survey for the respondents who were interviewed in those years.

  Restricting the sample to those present for all three waves from 2006-2010 reduces the total sample size by 34% (but by approximately 27% for individuals who have data on our main outcomes). The results are similar using the entire sample of respondents (see Appendix Tables A2 and A3, columns 1 and 2). However, due to missingness in the controls variables, the sample that remains when including the full set of controls is actually the almost identical to the analytical sample. As there is considerable item non-response, column 4 in Appendix

[2] This survey has also been referred to as the Malawi Diffusional and Ideational Change Project (MDICP) in the past. With major data collection rounds in 1998, 2001, 2004, 2006, 2008, 2010, and 2012 for up to 4,000 individuals, as well as ancillary surveys and qualitative studies, the MLSFH has been a widely-used dataset for research on health, family dynamics, social networks, and HIV infection risks in a rural SSA context (Kohler et al. 2015).

[3] GPS coordinates were obtained for respondents who took the HIV test in 2004, which is 91% of respondents (see Thornton (2008)), of which another 7% did not have valid GPS readings.

[4] Of these individuals, 1871 had non-missing basic demographic characteristics such as age, gender, and education, but only 1510 had completed the subjective expectations module, reported savings and expenditures–our outcomes of interest.

Tables A2 and A3 show the results for the restricted sample that have all the data for all years the outcome is available (such that the sample is fixed across outcomes).

The section on attrition (B) investigates the degree to which sample attrition affects our results. To extend our analysis, we incorporate data on a migration followup that occurred in 2013 (Anglewicz et al., 2015). The migration followup study was able to location and interview 246 of our starting sample that were lost to followup. The migration analysis, described in detail in B includes the 2013 wave of data for these migrants, analogous to an "Intent to Treat" analysis.

- *Treatment of outliers:* For monetary outcomes such as savings, expenditures on children, self, and farm inputs there are many zeros as well as many extreme values. For all models in the analysis in the main text, we transform all nominal values using the inverse hyperbolic sine transformation.

  The results are qualitatively similar using different approaches to handle zeros and outliers. Previous versions of the paper winsorized at the 98th percentile, and we include all main results using this specification in the appendix (see Appendix Tables A15-A20).[5] Our results are robust to specifications using the inverse hyperbolic sine and log transformations (and robust to using a range of values for the added constant), and as well as other thresholds for the winsorization (e.g. 99th, 95th, etc). Our results are also qualitatively similar using quasi maximum likelihood poisson estimation with fixed effects, though the estimation is frequently not possible due to known convergence issues when adding in controls as maximum likelihood estimates do not exist (Santos Silva and Tenreyro, 2011, 2010). Finally, we also show results for a Quantile Treatment Effect analysis using changes in outcomes from 2006-2010, and defining the binary treatment indicator for individuals near (within 6 km) of the ART facility in 2008. This is similar to a quantile difference-in-difference approach. The results are shown graphically in Figure A9.

A.2 MLSFH Child Sample

For the child sample, we use data constructed from household rosters (the linkage process was done by Castro et al. (2015)). The MLSFH household rosters do not provide unique identifiers for each person who is listed by the respondent on the household roster that was collected during the 2006-10 MLSFH surveys. To allow for longitudinal analyses of the information elicited in these household rosters, the data on the respondent's children listed in the 2006, 2008 and 2010 MLSFH household rosters were linked using names, ages, sex, and birth order. Because not all data were available in every wave, and because the spelling of names is not always exactly identical across waves, the matching was not undertaken with a computerized algorithm, but was done case-by-case instead.

[5] For earnings and total expenditures, fewer than 5% of respondents report null values, and so we always transform these variables using the inverse hyperbolic sine (and these results are also not sensitive to the transformation used).

Two processes were undertaken simultaneously. First, names were designated the principal matching variable; so to be consider matched, a minimum similarity in spelling was required. Second, a quality indicator for the quality of the match was assigned to each matched child, with the match being *low quality*, if no other data than the spelling itself was available to establish the match, and the spelling itself was not sufficiently similar across waves, *medium quality*, if any other variable was available (age, sex, birth order) to establish the match or, if no other data was available but the spelling matched very closely, and *high quality*, if two or more variables were available to establish the match. Only children of medium and high quality of match are included for the analyses of this paper, which represent about 90% of the total matched cases. See Castro et al. (2015) for additional detail of this linkage process.

We restrict the analysis on the subsample of biological children living with the respondent that are between the ages of 5 and 19, inclusive, as children not yet eligible to start school or who are already too old would not be affected by the rollout. Like in the adult sample, we include children who are observed in the data for all waves of the survey.

A.3 Spatial Data

All spatial analysis was performed using ArcGIS. Distance calculations were computed using ArcGIS Network Analyst, which uses Dijkstra's algorithm for finding the shortest paths on an undirected, nonnegative, weighted graph.

- *Clinic Data, HIV Unit at the Malawi Ministry of Health:* We obtained data on ART start dates, GPS coordinates, and other clinic characteristics from the HIV Unit at the Malawi Ministry of Health. Clinic database with GPS coordinates did not contain ART service start dates, but a separate list of clinics providing ART with start dates was obtained from the HIV Unit and matched by hand to the clinic database.

- *Road Network Data, Malawi National Statistics Office (NSO):* In order to determine distance by road, we incorporate road network data (arcGIS files) provided by the Malawi National Statistics Office (NSO). The NSO data also contained coordinates of important landmarks such as schools and trading centers.

- *Population Density, Afripop data:* We use estimated population density at a resolution of 100 meters from the Afripop database, which is now a part of the WorldPop project (Tatem et al., 2007; Tatem and Linard, 2011). Population density is estimated using census data along with satellite imagery for mapping settlements (30m spatial resolution Landsat Enhanced Thematic Mapper), in addition to other sources.

B Attrition Analysis

Appendix Table A10 provides summary statistics by attrition status and shows significant differences between attritors and non-attritors. Attritors tended to be younger, male, unmarried, more educated, closer to major roads, and more likely to be HIV-positive. As described in the data section above, we estimate Equation **??** using a balanced panel. Thus our results are internally valid (even if there is attrition) as long as attrition is not correlated with ART proximity. We worry that individuals who live far from ART and are forward-looking may move closer to ART facility (and attrit from the sample). Thus the sample of respondents that remain in the analysis would include both types (forward-looking and not) in the near areas and only present-biased types in the far areas. Several pieces of evidence suggest this is not the case. First, attrition is not correlated with ART proximity, which can also be seen in Table A10 as the "Distance to ART in 2008" (Panel E) in columns 1 and 2 are not statistically different (if anything attritors were closer to ART facilities, p=0.11, but the difference in ART *proximity* (ln-distance) has a p-value of 0.25). Second, attritors did not appear to be less present biased, there is no difference in whether the respondent exhibited a "High discount rate" (Panel B) by attrition status.

Appendix Table A10 also repeats the balancing tests (regressing characteristics on ART proximity controlling for spatial characteristics) for the attritor sample. Among attritors, people near ART were more likely to be HIV-positive and report more sick relatives that died of AIDS. Attritors near ART also had more education. And they were more likely to have a mobile phone, a pattern that also appeared with the non-attritors. However, there was balance in the main outcomes of interest (Panel D) among attritors, and overall characteristics are not jointly correlated with ART Proximity, with a p-value of the joint test equal to 0.53.

Finally, we present estimates of our main results on savings and child expenditures using Inverse Probability Weights (IPW) to account for attrition. IPW will correct for a bias in the estimated treatment effect under the assumption of attrition on observables, z (and where z is a larger set of observables than those used as controls in our main estimation procedure. In other words, let $s_i = 1$ if a respondent is observed in 2006 and does not attrit from the sample in later waves, let z be the set of observables, and y the outcome in 2010. Selection on observables implies that $P(s = 1|y, z) = P(s = 1|z) = P(z)$, that is conditional on observables the probability of attrition is independent of the outcome. In this case, we use the observables in 2006 to predict (using a logisitic regression model) whether a respondent is in the sample to get an estimate of $P(z)$, $\hat{p}(z_i)$. We then weight the observations in the balanced panel regression using $1/\hat{p}(z_i)$ for sampling weights. We further allow the observables to affect attrition differently depending on x, the ART proximity (here we define x as a dummy variable equal to 1 if near (<6km from) ART). Thus, we separately estimate $\hat{p}(z_i|x_i = 1)$ and $\hat{p}(z_i|x_i = 0)$. The results of the main analysis redone with IPW are presented in Appendix Tables A12 and A13. The IPW estimates similar to those reported in the main text, but the coefficients for the savings results are approximately 5-10% smaller, while the coefficients on education expenditures are approximately 20% larger. The coefficient on other spending also increases, and the increase in medical expenditures are marginally significant though no in all specifications, while spending on clothing still remains unaffected by ART (although point estimates are more positive, we cannot reject that they are different from zero).

Next, we turn to incorporating attritors who were located in the 2010 wave (183 cases) as well as those who were found at the 2013 migration followup study (246 cases). Appendix Table A11 shows characteristics of the analytical sample, attritors located (in 2010 or 2013), and attritors who were never located (287 cases). We will refer to attritors located in 2010/2013 as migrants, and those never located as lost to followup (LTFU). Note that the characteristics of migrants and LTFU differ along different dimensions to the analytical sample: migrants were younger, had fewer children, and higher education while LTFU were living in more densely populated areas. Both migrants and LTFU were more likely to have HIV, perceive higher mortality risk and HIV threat, and live closer to major roads and clinics. Importantly, ART proximity in 2008 was not different across the three groups.

As a further robustness exercise to test the degree to which attrition can explain our results, we include migrants back into our analysis, in an "Intention to Treat" approach. This increases the sample size to 1808 individuals.[6] We have 2010 values for migrants who were located in 2010, however, we use 2013 values for outcomes for the migrants located in 2013. However, migrants found in 2013 were not asked expenditures questions, so we focus the analysis on the savings and mortality risk outcomes.[7] The results of our main regressions including the migrant sample are in Appendix Table A14, Columns 1,4, and 7 reproduce the results from the analysis in the main text. Columns 2,5, and 8 include the migrants. Finally, Columns 3, 6, and 9 estimate the previous regression with sampling weights to adjust for remaining attrition. We again use Inverse Probability Weighting, calculating new weights for each respondent for each wave of the survey from 2008, 2010, and 2013, ie, the sampling weight, $\frac{1}{\hat{p}(z_{it})}$, now varies over time within the panel (see Wooldridge (2007)). Our results are robust to including temporary attritors and migrants, though the point estimates on savings decrease somewhat (by approximately 15% across different years) when including attritors. The point estimates on the mortality risk index increase when including attritors. Estimation using sampling weights (IPW) yields very small changes to the estimated coefficients.

[6] Appendix Table A11 Columns 4-5 test "balance" along the ART Proximity gradient, using the analytical+migrant sample, and the results are quite similar to the analytical sample only (metal roof, mobile phone and mental health are still significantly correlated to ART Proximity while there is balance along the main outcomes).

[7] Only a subset of mortality expectation were asked for the migrants: the 1, 5, and 10 year own-mortality probabilities. Thus the mortality index, which is a factor score of all mortality perceptions, is calculated using these 3 measures.

C Imputing Savings

The MLSFH survey did not begin collecting data on savings until 2006, and so we are unable to directly test for pre-trends in savings. In order to assess the potential that our savings results are driven by a pre-existing trends, we model savings in 2006 in order to impute savings in waves prior to 2006. For this exercise, we incorporate data from the 1998 and 2001 waves in order to obtain a longer time horizon to estimate trends in savings.

We estimate (log) savings in 2006 using a rich set of demographic, economic, and HIV-related variables and then use the parameters of the model to impute savings using the available data collected in previous waves when savings itself was not collected. This exercise is similar to testing pre-trends across individual variables, as we have done in Table **??**; however, it combines the available variables in a way that best correlates with the outcome of interest. While this test is stronger than testing pre-trends of individual variables, it cannot rule out the possibility of pre-trends in savings that are uncorrelated with observable characteristics.

Formally, we estimate

$$\ln(savings)_{2006} = X_{2006}\beta + Z\gamma + \varepsilon_{2006} \tag{A1}$$

where $\ln(savings)$ is the vector of savings (transformed using the inverse-hyperbolic sine) for all individuals in our analytical sample in year 2006, X_{2006} is a matrix consisting of the subset of predictors or covariates of savings in 2006 (and Z includes the standard set of spatial controls used in the main text which do not vary over time), and β is the vector of parameter estimates. We then obtain $\hat{\beta}_{2006}$, and predict $\ln(\widehat{savings})_j = X_j\hat{\beta}_{2006}$ for each $j \in \{1998, 2001, 2004\}$.[8]

Technically, the set of covariates available in each survey wave is not the same. For example, in 2004 we have a measure of the individuals' rate of time preference, while in 2001 and 1998, we do not. On the other hand, earnings were not reported in 2004, but they were in 2001 and 1998. To account for this, we estimate the base model in equation A1 using 3 sets of covariates: each set is reflective of the covariates that are actually available in the previous waves of the survey. This ensures that the coefficient accurately reflects each covariates' contribution to predicting savings, conditional on the other available covariates.[9] Note, the parameter estimates, β^j_{2006}, are always obtained using 2006 covariates.

Table A7 presents the model used to impute savings using all available covariates (as mentioned above, we estimate this model 3 times, using only those covariates in 2006 that were also available in the survey wave we are imputing for). The set of demographic, income, expenditure, and HIV-related variables included in the model are all reported in the table. The coefficients on the spatial

[8]We have also modeled total savings using a Tobit model as well as the binary savings decision using a Probit specification and found similar results in both cases.

[9]Formally, we use X^j_{2006}, which is a matrix consisting of the subset of predictors in 2006 that are also available in year $j \in \{1998, 2001, 2004\}$, to obtain estimates of β^j_{2006} which is the vector of parameter estimates using the subset of predictors available in year j. We then obtain $\hat{\beta}^j_{2006}$ for each $j \in \{1998, 2001, 2004\}$, and predict $\ln(\widehat{savings})_j = X^j_j\hat{\beta}^j_{2006}$ for each $j \in \{1998, 2001, 2004\}$.

controls have been suppressed to save space. We note that coefficients are of the expected sign: married and wealthier individuals have more savings, while individuals with more children or high discount rates have fewer savings. Ownership of certain assets, particularly beds, goats, and metal roofs, are strong predictors of savings. Earnings are also a strong predictor of savings, and the elasticity of the stock of savings with respect to earnings is about 0.14. Furthermore, individuals who worry about HIV or are HIV positive tend to have higher savings, indicating that the AIDS risk is a potential motivator for precautionary savings. Our model explains 19% of the variation in 2006 cash savings.[10]

To ensure our imputed savings have predictive power, we obtain the R^2 from regressing actual measures of savings on imputed savings in 2008 and 2010 (when we observe both measured and imputed values).[11] Table A8 reports the share of the variation in total savings explained by the imputed savings in 2008 and 2010: our imputed measure explains 22% of the variation in savings in 2008 and 29% in 2010: a considerable amount of variation. Figure A5a shows local linear regression, with 95% confidence bands, of actual savings as a function of predicted savings in the three waves both were available: the functions lie close to the 45-degree line (with evidence of censoring, suggesting a Tobit model would perhaps be more appropriate for estimating A1).[12]

We also test if the imputing savings variable generates treatment effects in 2006-2010. Given the set of covariates used to model savings, only "Worried about AIDS" actually reflects beliefs about mortality risk. As a result, we should not expect imputed savings to respond strongly to ART proximity as the co-variates we use to impute savings should not change with shocks to beliefs alone. Table A8 shows that indeed, the imputed savings (column 1) do not respond to ART (column 2 shows the actual response in savings using the same sample).[13]

Table A9 estimates the pre-ART trends in imputed savings, as well as three important variables: total earnings, roof structure (a proxy for wealth), and whether the respondent was worried about AIDS. The dependent variable in Column 1 uses imputed log-savings using variables for years 1998-2004, but actual reported cash savings from 2006. The coefficient on $Year \times ARTProx$, is very small and insignificant. In fact, the pre-ART trend for savings is actually slightly negative, and the standard errors are also fairly small indicating our null findings on trends are not due lack of power.

We also show results for pre-trends for 3 variables that we measure from 1998 onwards and that are highly correlated with either savings or mortality risk: earnings, having a metal roof, and worry about AIDS. We chose these variables because they respond to economic and diseases

[10] 18% using just the subset of variables available in 2004, and 17% of the variation using the subset of variables available in 1998 and 2001 waves.

[11] For 2006, we have already reported above that 19% of 2006 savings are explained by covariates, but the share of variation explained in 2008 and 2010 are particularly of interest because they are out-of-sample predictions.

[12] The results are similar using a Tobit model for savings, though for simplicity, we show all results here using a linear model.

[13] Including mortality risk variables into the set of predictors of savings in 2006, however, does yield significant (though, substantially smaller) effects of ART proximity on imputed savings: in 2010, the coefficient on ART proximity is 0.14 ($p<0.05$). Since we do not have these variables before 2006, we cannot include them in the imputation for savings prior to 2006.

environment changes over time and not just fixed demographics. The coefficients in Table A7 mask the importance of earnings and roof structure in predicting savings because of the many additional covariates. However, regressing savings in 2006 on earnings, metal roof, and worried about HIV reveals an elasticity of savings with respect to earnings of 0.23 ($p<0.001$) and a semi-elasticity of savings with respect of having a metal roof of 0.74 ($p<0.001$), while worried about AIDS remains uncorrelated with savings.[14] Columns 2-4 of Table A9 estimate the pre-ART trends for earnings (transformed with the inverse hyperbolic sine, though earnings were not reported in 2004), roof structure, and worry about AIDS. The coefficients on $Year \times ARTProx$, are very small and insignificant, though fairly precisely estimated across all 3 outcomes, indicating no differential pre-trends in economic outcomes or disease environment.

Finally, Table A9 Column 5 estimates the main savings results with the addition of a control for the pre-ART annual growth rate in savings interacted with year, which reduces the magnitude of the point estimate in 2008 by 20 percent (from 0.69 to 0.43, no longer statistically different from zero), and in 2010 by only 6 percent (and still highly statistically significant).[15]

[14]We nevertheless show this outcome because it is strongly associated with own mortality risk in 2006: a similar regression of own 5-year mortality risk on earnings, metal roof, and being worried about AIDS reveals that while earnings and metal roof are not significantly correlated mortality risk, being worried is associated with a 0.08 greater 5-year mortality probability ($p<0.001$).

[15]As in the main text, column (5) sample is restricted to individuals who were interviewed in all three years of the survey (2006, 2008, and 2010).

D Effect of ART on Subjective Life Expectancy

Table A4 shows the effect of ART on components of the mortality risk index. While we hesitate in drawing any strong conclusions from each individual point estimate (because of noise in the dependent variable and concerns over multiple inference), we discuss some of the broad patterns. Own and healthy mortality risk behave similarly (except own 10-year mortality risk) and respond to ART by 2010 but the estimates are imprecise. Meanwhile, HIV+ mortality risk responds the strongest to ART proximity across all time horizons (column 3) and is statistically evident by 2008. This is reassuring as the HIV+ mortality risk precisely the variable that should respond to ART availability (since own and healthy mortality risk are affected by compounding the probability of getting HIV and then mortality risk conditional on positive). Own mortality risk adjusts most strongly and earlier for the subset of respondents who report any likelihood of HIV at baseline (column 5). Finally, although people in Malawi generally believed ART to be extremely effective early in the rollout, it's unlikely that they would have known the long-term efficacy of the medication (indeed, in the developing world context, this was even unknown to scientists at the time). As such, our results are consistent with a model of learning about the long-term efficacy of ART (column 4): only the 10-year horizon for ART mortality risk is affected by ART proximity and this is only evident by 2010 (consistent with learning taking time, as people observe those on the medications still living).

By contrast to results above, Panel C, Columns 1 and 2 indicate that 1-year own and hypothetically healthy mortality risks also respond to ART availability. However, 1-year mortality risk results not medically realistic given the long period of asymptomatic infection before the disease progresses to AIDS. However, respondents' perceived mortality probabilities of hypothetically healthy individuals are likely to include future risk of infection. Furthermore, respondents may not have accurate information on disease progression, the future infection risk is incorporated even at 1-year horizons. Respondents may have incomplete/inaccurate knowledge about the technical aspects of the disease because information about HIV/AIDS was not discussed in the public sphere until 2001.[16] Thus, respondents may form beliefs largely based on what they observe about the disease, which is usually fully developed AIDS. Additionally, respondents overstate other risks associated with HIV/AIDS, a finding that has been well-documented in numerous studies with independent samples across Malawi (Kerwin, 2016; Derksen and Muula, 2016). For example, in our data the reported transmission probability per sex act is close to 90%, whereas the actual transmission probability is less than 1%. These patterns may be explained by a health education policy which purposely overstates the risk of contracting HIV in an effort to discourage risky behavior. Such policies, aimed at "scaring" individuals to induce behavior change, did not always convey the nuances of the disease (Barden-O'Fallon et al., 2004). For example, only 65% of women in 2004 knew it was possible that healthy

[16] At the start of the epidemic in Malawi, President Banda believed that issues relating to sex, including HIV transmission, should not be addressed in the public sphere. Until 1994, it was illegal for Malawian citizens to discuss the epidemic openly, and even after the change of power to President Muluzi, it was not until 2001 that information about HIV/AIDS became available via TV and radio (Kalipeni and Ghosh, 2007).

looking people can have HIV (Barden-O'Fallon et al., 2004), and that rate was lower among more rural and less educated women. As a result, many people associated HIV infection with the symptomatic manifestation of the disease (by the time symptoms are visible, the individual has AIDS, and life expectancy without ART is less than 1 year). Thus, while medically inaccurate, the limited understanding of the nuances of the disease combined with a policy that purposely overstated risks may explain why we see results at the 1-year horizon even for hypothetically healthy individuals.

Two pieces of evidence in our data suggest the results on healthy mortality risk are driven by HIV-related risk: (1) Perceived current and future HIV prevalence are both highly correlated with healthy mortality risk at all horizons (see Table 1 below), but are consistent with timing in that current HIV prevalence is more important for the 1-year horizon while future prevalence is more important for the 10-year horizon, and (2) partialling out the HIV+ mortality risk eliminates the effect of ART proximity on healthy mortality risk. Table A5 shows that after partialling out the effects of perceived HIV+ mortality risk, the coefficient on ART Proximity is substantially smaller and no longer marginally significant.

Table 1 – Correlations between perceived HIV prevalence and health mortality risk

	(1) 1-year healthy mortality risk	(2) 5-year healthy mortality risk	(3) 10-year healthy mortality risk
Current HIV prevalence	0.083***	0.065***	0.028
	(0.020)	(0.024)	(0.027)
HIV prevalence in 5 years	0.012	0.046***	0.080***
	(0.014)	(0.017)	(0.022)

* $p < .10$, ** $p < .05$, *** $p < .01$
Note: Standard errors (in parentheses) are clustered by village. All probabilities (HIV prevalences and mortality risks) are re[ported between 0 and 1 at 0.1 increments. Survey waves used are 2006, 2008, and 2010, and regressions control for region-by-year fixed effects.

Next, we calculate the implied effect on subjective life expectancy (SLE) from the estimates on 5-year mortality risk. We use the estimates reported in Baranov et al. (2015), Table 8 which report the results for the subsample of HIV-negative and non-caretakers (NCT).[17] We use the 5-year mortality risk measures specifically because it corresponds to the 5-year age groupings in mortality tables, and also because we worry that respondents hesitate to project their own mortality risk so far into the future. Interpreting the estimates of the impact of ART availability on subjective mortality risk as the first-stage is a valid approach if our identification strategy is only picking up the effect expectations and not direct effects of mortality changes, and the exclusion restriction is more likely to be satisfied among the HIV-negative/NCT sample.

The implied subjective life expectancy is calculated by adjusting the age-specific mortality rates

[17] The point estimate in Table 8, Column 2 is 0.02, which is the effect of going from 8 to 4km. The corresponding effect of going from 9 to 3.2km (the 5.8km difference we use in this paper) implies a decline in 5-year mortality risk by 0.04, or 4 percentage points. Our own estimates of ART proximity on the 5-year mortality risk in Table A4 are nearly identical, at 0.042 percentage points.

from life tables,[18] which provide age-specific mortality rates in 5-year increments (e.g., mortality rate for the age group 20-24). To estimate the implied life expectancy gain, we adjusted the age-specific mortality rates between the ages of 22 and 57 (this is generally the age range for the majority of AIDS-related mortality).[19]

We assume that the level change in subjective mortality risk reported using beans is an accurate reflection of the respondents' risk assessment. Thus, we apply the 4 percentage point decrease to the relevant age categories,[20] which yields a subjective life expectancy gain due to moving ART 5.8km closer to the respondent of 6.0 years.[21]

We can compare these subjective life expectancy gains to the estimated life expectancy gains of eliminating AIDS in Malawi (UN World Population Prospects, 2010 Revision). According to these estimates, eliminating all AIDS mortality in Malawi would lead to an increase in observed life expectancy at age 5 of 12.7 years. The gains provide a reasonable upper bound for the effect of ART on subjective life expectancy, since ART medication does not entirely eliminate AIDS mortality.[22] In light of the large change in objective risk, the strong response in subjective mortality risk and other behavior changes among the HIV-negative sample does not seem implausible.

[18] We use life tables from the UN Population Division for Malawi in 2009 since these are calculated based on mortality data the 5 years prior, which corresponds best to the years of our survey.

[19] The estimated effect on 5-year mortality risk from our respondents is an average based on all respondents and is not age-specific. While it would be possible to estimate the effect of ART on 5-year mortality risk for each age group, the estimates become less precise. Thus we use the overall effect of ART on 5-year mortality risk to adjust the age-specific mortality rates in the age groups where AIDS mortality is occurring. In practice, if we take the age-specific mortality reductions at face value, our approach underestimates the life expectancy gains because the youngest cohorts report the largest mortality risk declines.

[20] Except for the 20-24 age group, which we treat separately, since that would result in negative mortality probabilities. Here we smooth the values by interpolating the decrease in mortality risk linearly between the age group above and below.

[21] An alternative approach assumes instead that the percentage change in the level is meaningful. Then 4 percentage points more accurately reflects a 10 *percent* decrease in mortality risk (since the average perceived mortality risk before ART was 0.4). This second approach may be more valid if we are concerned with interpreting the response in levels, given the levels of perceived mortality risk as compared to life-table based estimates of mortality risks (Delavande and Kohler, 2009) seem implausibly high. Applying a 10 percent decrease to the 5-year death probabilities to the relevant age categories in the life tables yields an increase in LE of 2.0 years.

[22] The life expectancy gains may seem high given the prevalence of HIV. It is useful to realize that the lifetime risk of getting HIV is also much higher than the prevalence, and in a country with 10 percent prevalence the lifetime risk of HIV is approximately 45 percent (Blacker and Zaba 1997).

E Savings and Growth

The effects of ART on saving behavior estimated in the main text used total cash savings as the outcome variable, which represents the total stock of liquid savings of the respondent at the time of the interview. In order to better interpret our results, we recast the savings results as the effect of ART on the personal saving rate. The MLSFH did not collect comprehensive data on expenditures, notably missing food expenditures and household items, so we do not take the standard approach to estimating savings as income less consumption. Instead, we take advantage of the panel structure of our data to generate savings rates as the change in cash savings between waves of the survey (net contributed to savings over 2 years) divided by income generated in those two years.[23] This approach yields a range in saving rates between -0.5 and 1 for 95% of the sample (we trim 2.5% of the sample from above and below to exclude nonsense values). The mean saving rate for the sample across both years is 12 percent (interestingly, the average national saving rate in Malawi over the period 2006-2010 was 14.7 percent).[24] On one hand, our measure of the individual saving rate will be biased upward because income reported does not include household food production that was consumed by the household. On the other hand, our measure of savings captures only liquid personal savings and does not include the value of land, farm equipment, or durable assets. We proceed, noting the caveat that our measure of the personal saving rate is capturing liquid savings as a share of market income.

As savings were not reported prior to 2006, we do not have pre-ART data on the saving rate. Our approach is no longer a double difference, and is only valid under the assumption that saving rates were uncorrelated with ART proximity prior to 2008. The combined facts that total savings and earnings were uncorrelated with ART proximity in 2006, and that trends in earnings and imputed savings prior to ART were parallel, suggest that this assumption is valid.

We regress the calculated saving rate on ART proximity, specifically, $s_{it} = \alpha + \beta \text{ART Prox}_i + X'_{irt}\gamma + \varepsilon_{it}$, where X_{irt} is the vector of spatial characteristics interacted with region and year, as in the main text and errors are clustered at the village level.[25]

Estimating the above specification, we find that the coefficient on ART proximity is 0.026 (with a standard error of 0.006) without additional spatial controls, and 0.038 (standard error of 0.010) once the full battery of spatial controls are introduced. The coefficient implies a reduction in distance of 5.8 km would increase the saving rate by 2.6-3.8 percentage points. Given the first stage effect of ART on implied life expectancy calculated in Section D, the corresponding life expectancy gains are 6.0 years. In other words, increasing life expectancy by one year implies an increase in the saving rate by 0.43-0.63 percentage points. Alternatively, it implies a 0.2-0.3pp increase in saving due to 1% (0.47 years) increase in LE (or an elasticity of the saving rate with respect to life expectancy of 2.8-4.1).

[23] We use earnings reported in year t of the survey, and estimate earnings at $t-1$ as the average between t an $t-2$ earnings. Note that like savings, income is also personal income and not household income.

[24] Statistics from the World Development Indicators, 2016.

[25] We use both 2008 and 2010 data, as both saving rates were calculated using "post" ART savings. We cannot use the fixed effects specification since ART proximity does not vary over time.

Next, we turn to investigating the plausibility of the magnitude of our savings results. We are, however, unaware of any other micro-based studies that produce an estimate of the elasticity of personal savings with respect to life expectancy. Instead, we turn to the cross country relationship between savings and life expectancy.[26] Figure A7 shows the relationship between average national savings between 1996-2006 and average life expectancy for that period.[27] There is indeed a positive relationship, with a slope indicating an increase in the saving rate of 0.38 percentage points for each year of life expectancy gained, though the R^2 is fairly low at 15%. Our estimates based on ART availability are approximately 13 to 66% higher than the cross country relationship. Alternatively, regressing the saving rate on log-life expectancy, we get an estimate of 0.19, ie an increase in LE by 1% is associated with an increase in the saving rate of 0.2pp.

We can also compare our estimates to that of Zhang and Zhang (2005), who model the effect of life expectancy on saving, schooling, fertility, and economic growth. Using cross-country data from 76 countries, the authors find that life expectancy has a significant positive effect on the saving rate: specifically, they estimate that a 1% increase in life expectancy increases the saving rate by approximately 0.24 percentage points (Table 2). Recall, based on the estimates in this paper, a 1% increase in life expectancy (an increase of 0.47 years) would lead to an increase of 0.2-0.3pp in the saving rate, very close the estimate from Zhang and Zhang (2005).

Our final exercise is to do a simple calculation to estimate the implications for economic growth, based on prevailing macro-literature on the relationship between savings and growth. Interestingly, the role of savings in growth has been questioned both empirically and theoretically, as countries can borrow to finance the purchase of capital and do not need to rely on domestic savings. However, a recent study by Aghion et al. (2016) argues that domestic savings can play a role in growth for poor countries because they enable local entrepreneurs who are familiar with local conditions to cooperate with foreign investors who are familiar with frontier technology. In such a model, growth results from innovations that allow local sectors to catch up with frontier technology and domestic savings acts as equity, mitigating an agency problem that would otherwise deter the foreign investor from participating.

Empirically, Aghion et al. (2016) do not calibrate their model, nor do they argue their estimates represent a causal interpretation. Nevertheless, we use their estimates as a benchmark, though major caveats apply. Based on the econometric specification implied by their model (the main implication is that there is no relationship between lagged savings and growth for countries close to the productivity frontier, but there should be a relationship for countries far from the frontier), an increase in the average saving rate of 10 percentage points is associated with an increase the growth of income per work by approximately 0.4pp (average growth over the 5-10 year horizon).

[26] We are very grateful to Hoyt Bleakley for this suggestion.

[27] While national savings are not equivalent to personal savings, it is reasonable to assume that personal savings comprise an important share of the national figure, especially in developing countries. Personal saving rates at the country level are only available for the 33 OECD nations and have a correlation coefficient of 0.3 with national saving rate. However, there is no relationship between either of the two saving rates and life expectancy for the OECD. This stylized fact is consistent with the results from Aghion et al. (2016), which we discuss in more detail shortly, who find that domestic savings are important for growth only among the relatively poor countries.

If the increase in saving rates due to ART were sustained (for 10 years, to be comparable with Aghion et al. (2016)), then combined with our saving results, a one year increase in LE would translate an increase in growth of approximately 0.017-0.025 percentage points in the long-run *due to the saving channel alone*.[28] Thus, our savings results imply that a 10 year gain in LE from eliminating adult AIDS mortality would lead to an increase in growth of income per worker by 1.7-2.5 percentage points.[29]

[28] Zhang and Zhang (2005) also provide an estimate of the effect of life expectancy on income growth: they find a 10% increase in LE is associated with a 0.09pp increase in growth, which is nearly an order of magnitude larger than the corresponding estimates based on our estimates combined with the Aghion et al. (2016) result. However, Zhang and Zhang (2005) estimates combine the effects of increased human capital and lower fertility in addition to increasing saving rates.

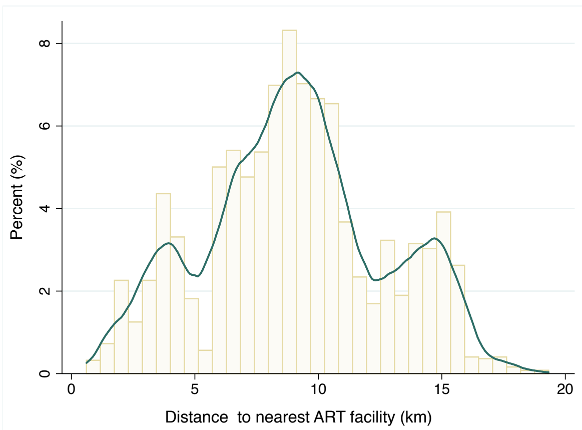
[29] Based on WHO projections using 2004-2009 mortality data, if ART eliminates all AIDS mortality in Malawi, it would lead to an increase LE by 12.7 years. Note the experienced increase in LE in the decade since ART became available is a staggering 14.5 years. The measured increase in life expectancy in the decade between 2004 (the start year of the large-scale rollout of ART, when life expectancy was 47) and 2013 is 14.5 years (World Development Indicators). Remarkably, Malawi's experienced life expectancy gains surpassed those based on estimates of entirely eliminating AIDS mortality. However, this comparison is misleading as the estimated LE gain from eliminating AIDS mortality use data based on mortality between 2004-2009, which actually already includes some effect of the ART rollout. To better compare the projection to measured gains, it makes sense to take the starting year for LE as 2007 (the mid-point for the mortality data used) until the latest available year, 2014. During that time, LE increased from 51.5 to 62.7, and 11.2 year gain.

The life expectancy gains cited above include child mortality, which is not relevant for the mechanism studied in this paper but is a considerable portion of LE growth during this time. Figure A6 plots LE conditional on age 5, based on WHO life table data. The change in age 5 LE between 2007 and 2014 is approximately 7.7 years, considerably lower than 11.2 gain which includes child mortality. For the purpose of estimating the impact of ART on growth, we focus on gains in adult (conditional on age 5) life expectancy: the gain from 2003 (just before ART) until 2015 has been approximately 10.8 years.

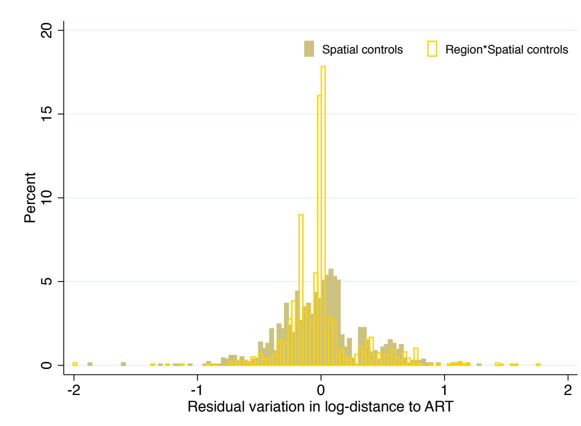
Appendix Figures and Tables

Figure A1 – Distance to ART Facility

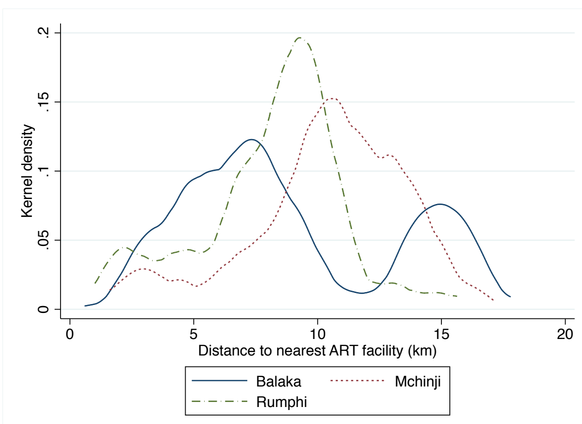
(a) Distribution of distances to ART in 2008



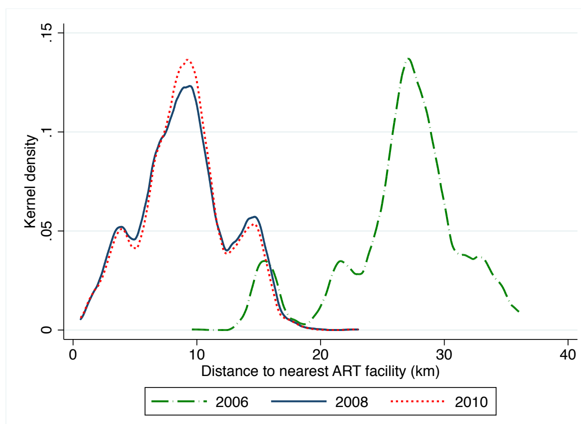
(b) Residual variation in ART Proximity



(c) Distribution of distances in 2008 by region

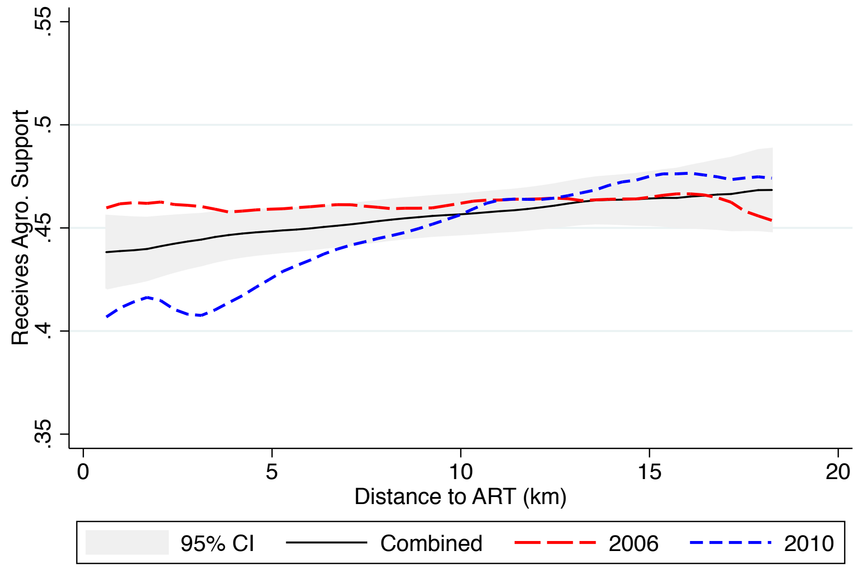


(d) Distribution of distances by year



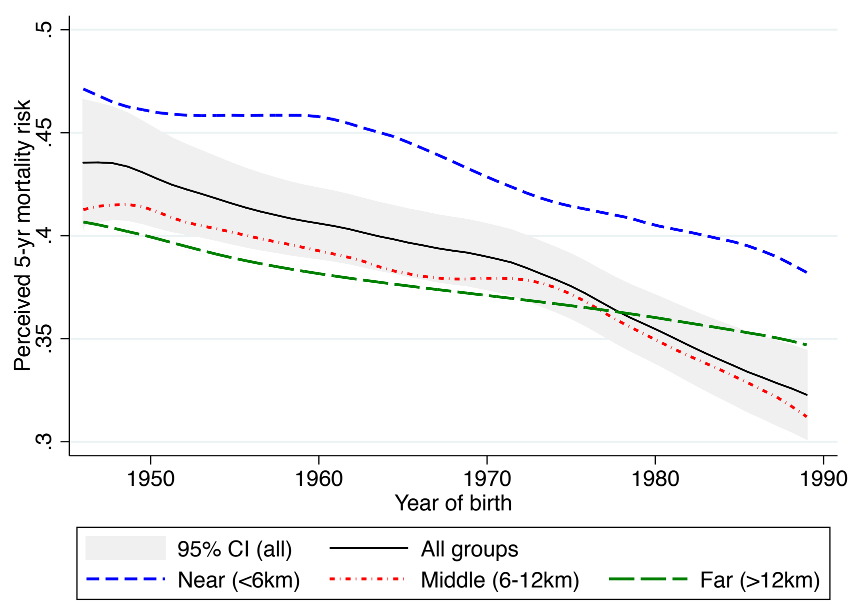
Notes: Panels A1a-A1c show the distribution of the main time-invariant explanatory variable: the distance to the nearest ART facility in 2008. Panel A1a combines all regions and shows the raw distribution in the background. Panel A1b shows the residual variation in ART Proximity (log of distance) after all spatial controls (interacted and not interacted with region) have been partialled out. Panel A1c show the distributions by region, noting that while the range and means are very similar, the overall patterns are not uniformly similar across the three regions. Throughout the analysis, we always control for region-by-year fixed effects in order to ensure that the patterns we observe are not due to the differential trends by region (which are unrelated to availability of ART). Panel A1d shows the distributions of distance to the nearest ART facility over time, for 2006-2010. As the ART rollout began in 2004, the nearest facility would be located in the major cities of Malawi only (greater than 200km for all respondents). By 2006, some district hospital were providing ART, though almost all respondents (99%) lived more than 15km away from their nearest facility. Between 2006 and 2008, clinics closer to the respondents began providing ART, shifting the distribution inward for all respondents. Between 2008 and 2010, new clinic openings did not largely effect the distance to the nearest ART facility for the respondents. For ease of interpretation, we use the time-invariant distance to ART in 2008 interacted with year to estimate the effects of ART availability. However, using the time-varying distance yields almost identical results.

Figure A2 – Farm Input Subsidy Program (FISP) participation and distance to ART year



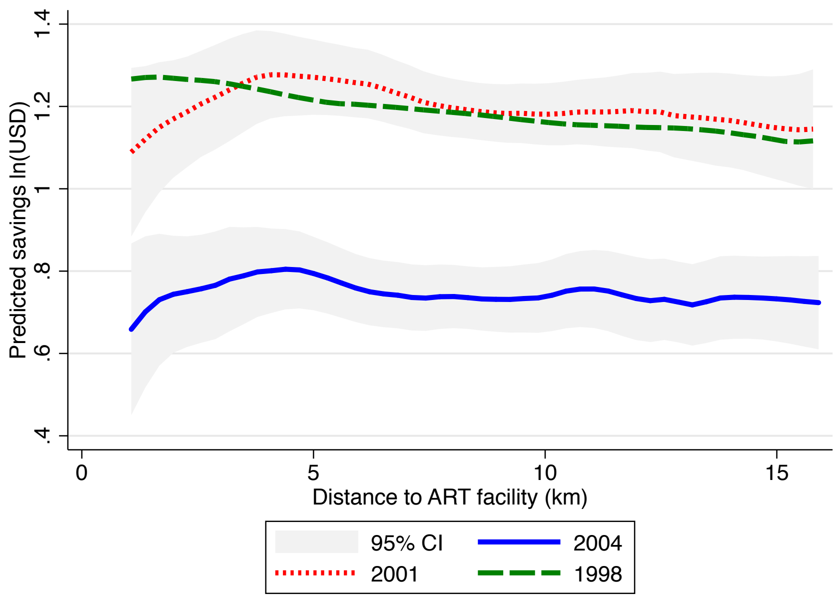
Notes: This figure shows the relationship between distance to ART and participation in the Farm Input Subsidy Program in 2006 and 2010 (no data exist in 2004 or 2008). As in the rest of the analysis, region-by-year effects have been partialled out. Overall participation (combined for both years) is also shown with a 95% confidence band. The figure indicates that households near ART were not more likely to receive agriculture subsidies.

Figure A3 – Subjective Mortality Risk in 2006 by Year of Birth



Notes: This figure plots local linear estimates of subjective mortality risk in 2006 as a function of birth cohort. Region effects have been patrialled out. Standard errors are not clustered or adjusted for the two-stage estimation, so they are only shown as guide. Since this graph represents a pre-trend test, standard errors that are downward biased represent a conservative approach, as we are more likely to reject the null, leading us to conclude that there were differential pre-trends. Even so, the results do not suggest that trends were diverging prior to ART.

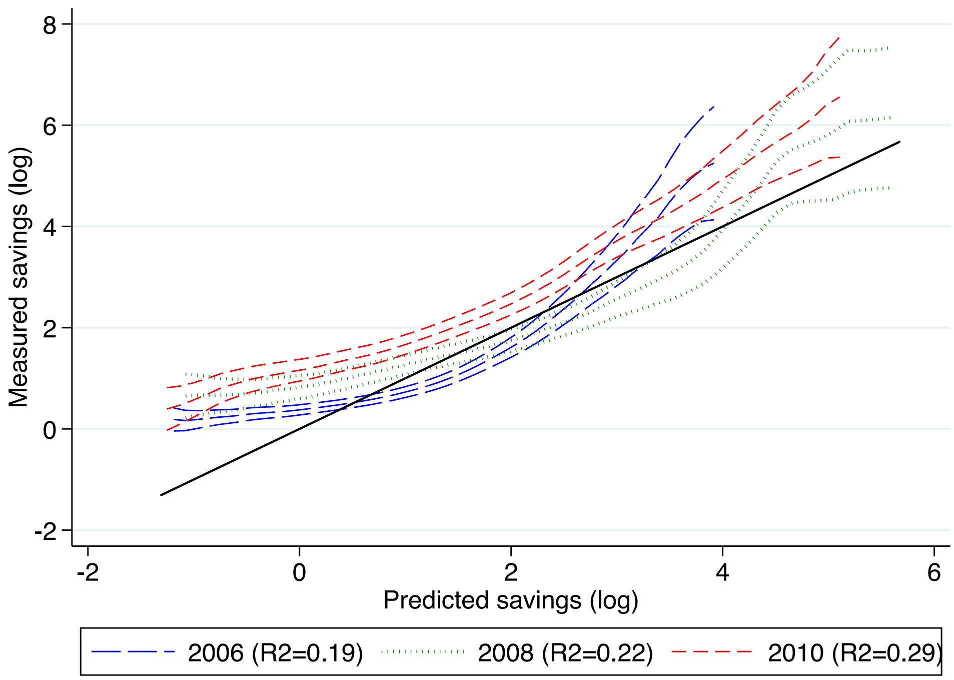
Figure A4 – Predicted savings as functions of distance to ART



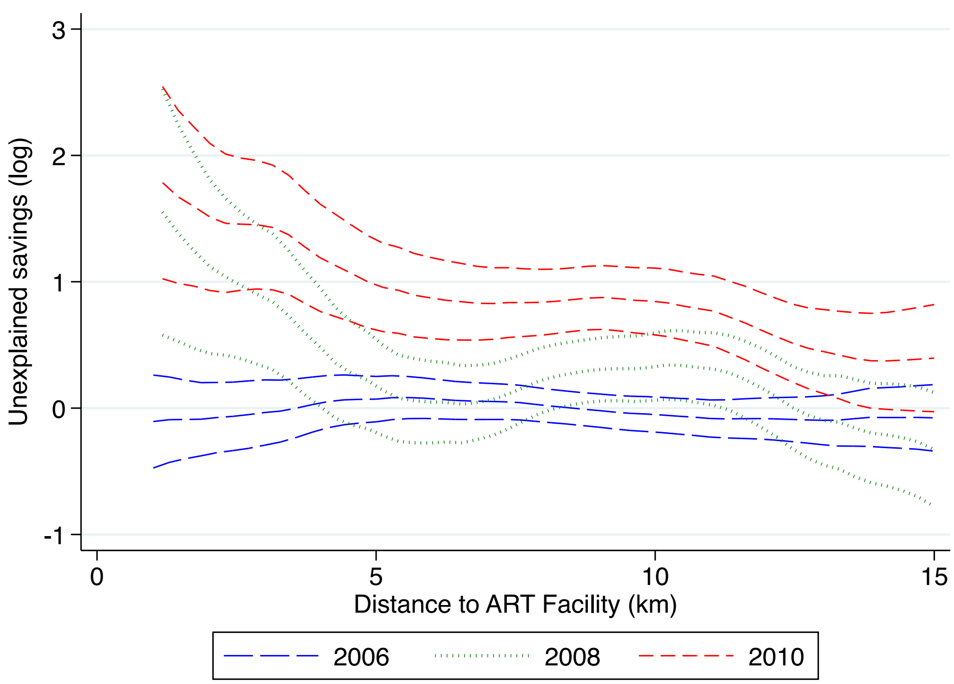
Notes: This figure plots local linear estimates of predicted savings (modeled as ln-savings, using the inverse-hyperbolic sine transformation) for respondents in the early waves of the MLSFH survey– for years 1998, 2001, and 2004 – as a function of distance to the ART facility. The confidence intervals are plotted for 2001 and 2004 years only (confidence band for 1998 is omitted for clarity). Savings are modeled using available socio-economic and demographic data in the relevant survey year. See Table A7 for the full list of variables used to predict savings.

Figure A5 – Predicted savings

(a) Actual savings as a function of predicted savings for each year

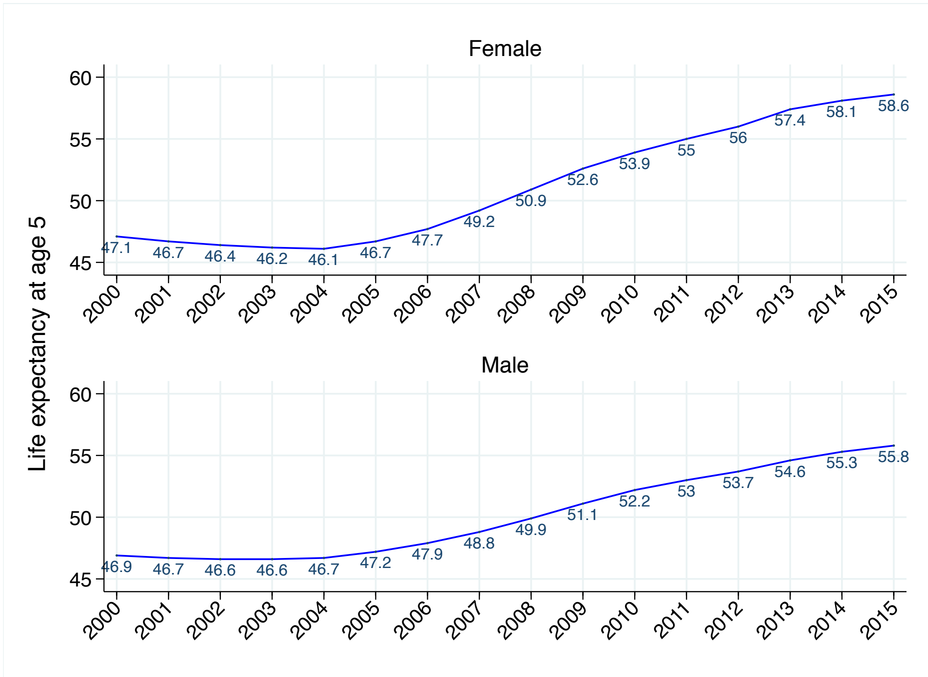


(b) Predicted savings as functions of distance to ART



Notes: Top panel of this figure plots local linear estimates with 95% confidence bands of actual savings as a function of predicted savings for each year. Bottom panel plots local linear estimates (with 95% confidence bands) of the unexplained variation in savings (actual–predicted) as a function of distance to ART for each year. Predicted savings are modeled using available socio-economic and demographic data in the relevant survey year and coefficients from the regression of savings on covariates in 2006. See Table **??** for the full list of variables used to predict savings.

Figure A6 – Life expectancy at age 5 in Malawi



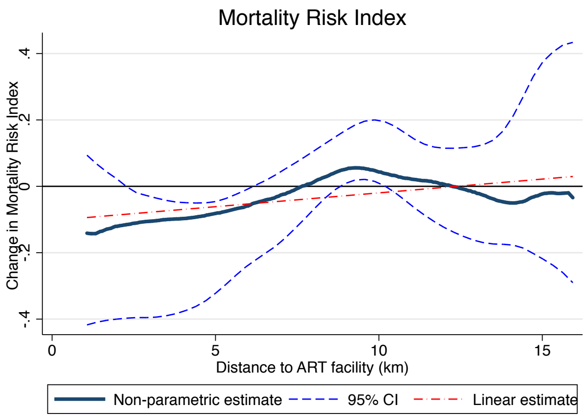
Notes: Life expectancy by gender for the age group 5-9 based on WHO life tables for Malawi (http://apps.who.int/gho/data/view.main.60980).

Figure A7 – Relationship between Saving and Life Expectancy in cross-country data

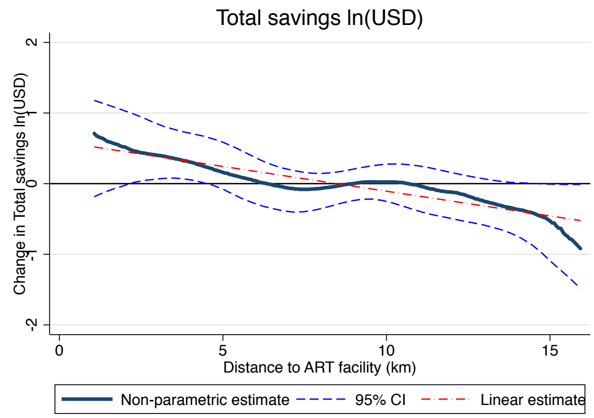
Notes: Figure uses data from World Development Indicators. The vertical axis is the average national saving rate over the 20 year period between 1996-2006, and the horizontal axis is the average life expectancy at birth over that same 20 year period.

Figure A8 – Non-parametric estimates of the effect of ART availability

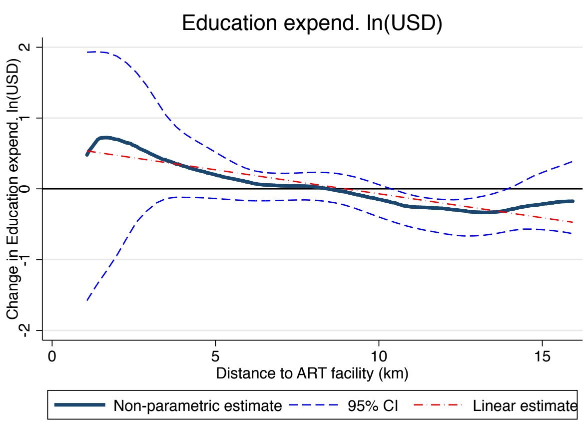
(a) Mortality risk



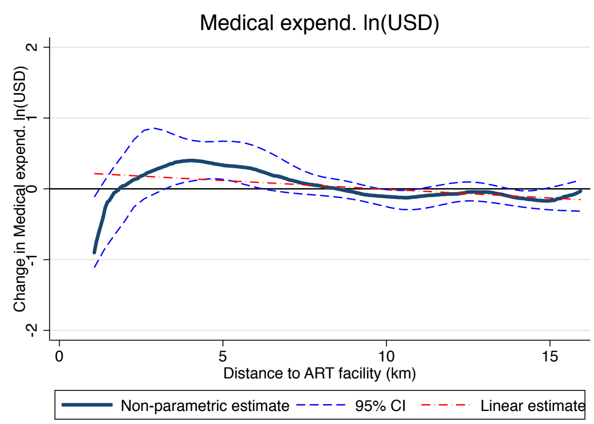
(b) Total savings



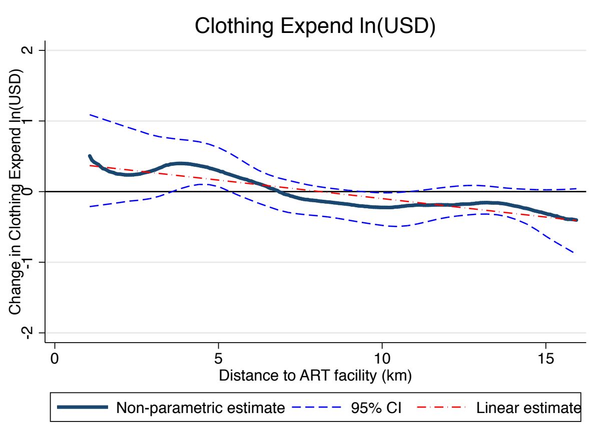
(c) Education expenditure



(d) Medical expenditure



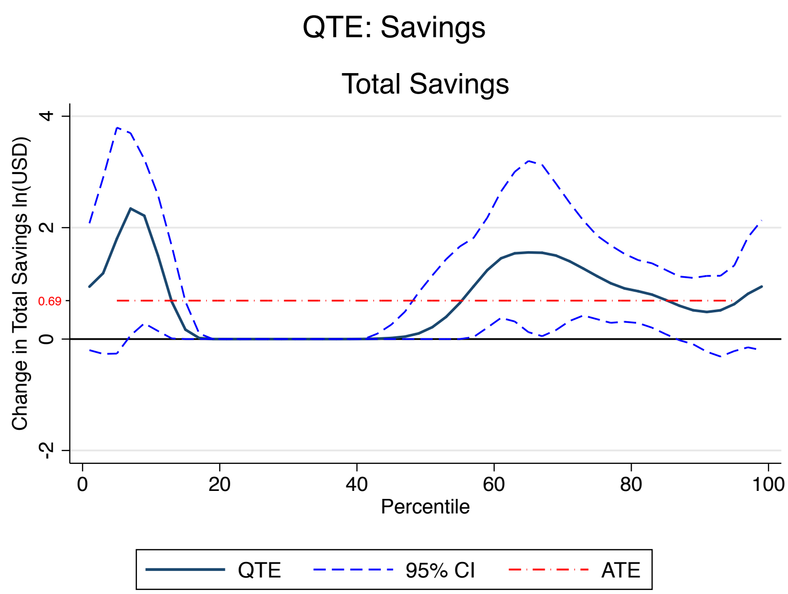
(e) Clothing expenditure



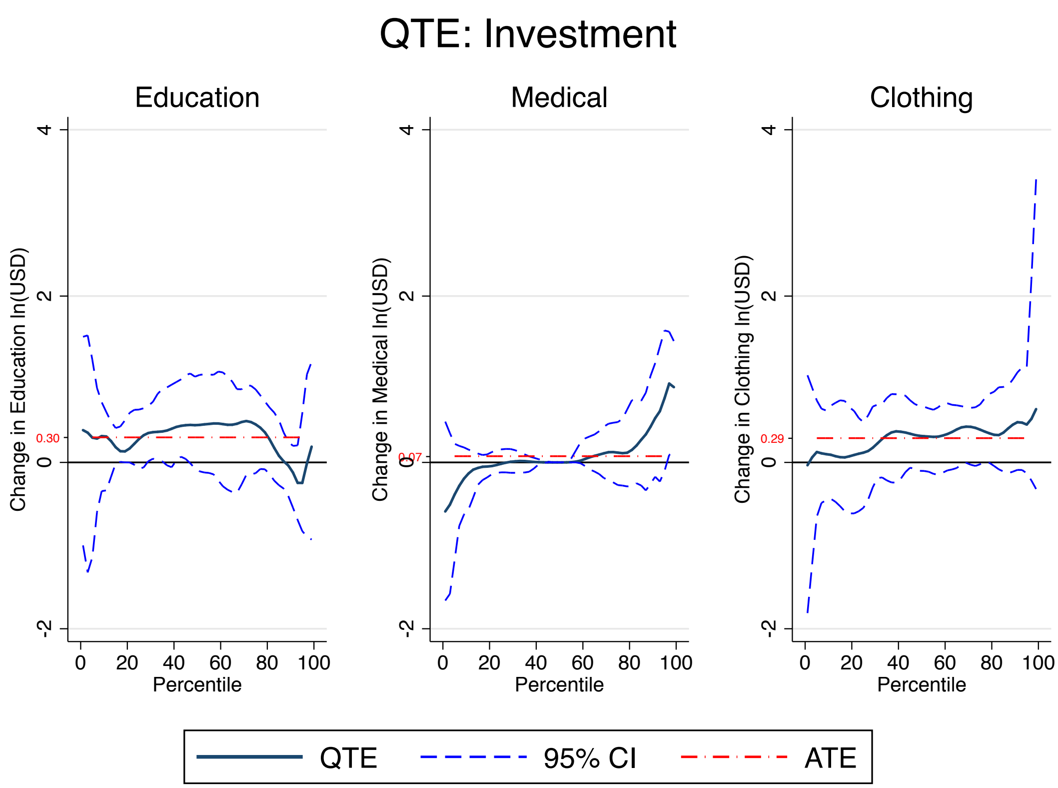
Notes: This figure shows nonparametric estimates of changes in the subjective mortality risk by distance to ART. Changes are computed using 2010 as the "post" year, and 2006 as the "pre" year. This corresponds to the coefficient on 2010 × *ARTProximity* in the parametric results presented in the tables. All graphs are nonparametric local linear regressions with region-year effects partialled out. Confidence bands at the 95% are computed using 1,000 bootstrap replications, clustered by village. In each bootstrap step, and undersmoothed local linear bandwidth is chosen following Hall (1992).

Figure A9 – Quantile Treatment Effects: Changes in Outcomes

(a) Saving Behavior



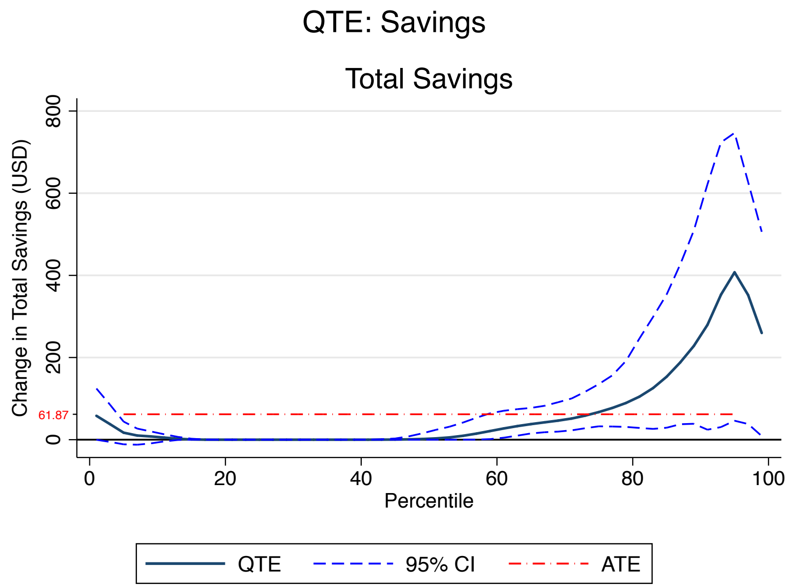
(b) Investment in Human Capital



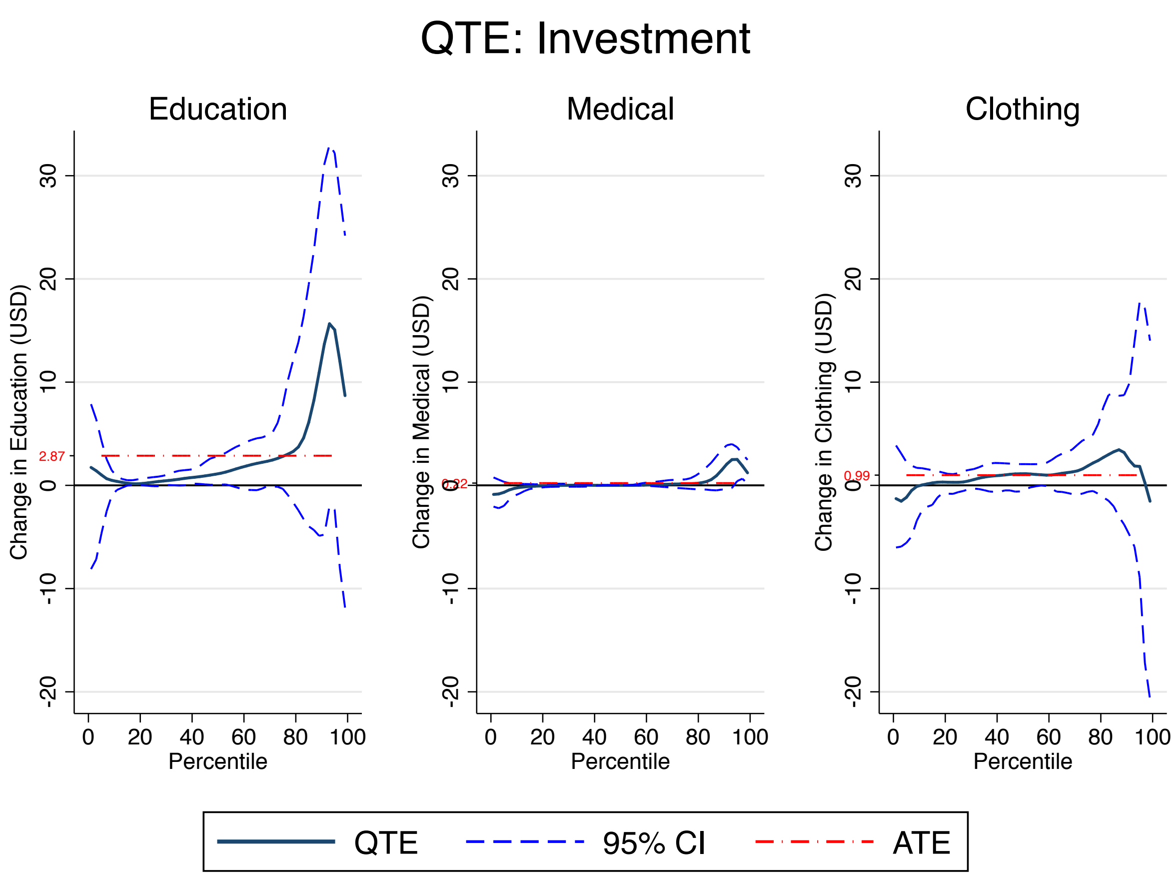
Notes: Quantile Treatment Effects of ART availability on *changes* in outcomes from 2006 (pre) to 2010 (post). Binary treatment variable is defined as treated = 1 if they are near (<6km) the ART facility, and zero otherwise. 95% confidence intervals for the QTE were calculated by bootstrapping using 1,000 replications with replacement, clustering at the village level. The average treatment effect (ATE), the mean difference in changes between near and far, is presented for comparison.

Figure A10 – Quantile Treatment Effects: Changes in Outcomes (Winsorized)

(a) Saving Behavior



(b) Investment in Human Capital



Notes: Quantile Treatment Effects of ART availability on *changes* in outcomes from 2006 (pre) to 2010 (post). Binary treatment variable is defined as treated = 1 if they are near (<6km) the ART facility, and zero otherwise. 95% confidence intervals for the QTE were calculated by bootstrapping using 1,000 replications with replacement, clustering at the village level. The average treatment effect (ATE), the mean difference in changes between near and far, is presented for comparison.

Table A1 – Characteristics of Clinics by ART Start Date

ART Start Date:	Before 2005 (1)	2005-2006 (2)	2007-2008 (3)	2009-2010 (4)	No ART (5)
Catchment Population	43709	54092	30453	22605	18972
	(25872)	(46095)	(14320)	(13901)	(14131)
Number of Beds	312.5	120.4	14.9	13.1	11.7
	(286.5)	(102.3)	(13.0)	(14.3)	(27.2)
Electricity	1.0	0.9	0.5	0.5	0.4
	(0.0)	(0.2)	(0.5)	(0.5)	(0.5)
Flush Toilet	1.0	0.9	0.4	0.4	0.3
	(0.0)	(0.3)	(0.5)	(0.5)	(0.5)
HIV Testing	1.0	0.9	0.9	0.9	0.8
	(0.0)	(0.2)	(0.2)	(0.2)	(0.4)
Outpatient	1.0	1.0	1.0	1.0	1.0
	(0.0)	(0.1)	(0.0)	(0.0)	(0.2)
Inpatient Maternity	1.0	0.9	1.0	0.9	0.7
	(0.0)	(0.2)	(0.0)	(0.3)	(0.5)
Inpatient General	1.0	0.8	0.2	0.2	0.1
	(0.0)	(0.4)	(0.4)	(0.4)	(0.3)
Antenatal Clinic	1.0	1.0	1.0	1.0	0.8
	(0.0)	(0.1)	(0.0)	(0.2)	(0.4)
STI Clinic	0.9	0.8	0.5	0.5	0.3
	(0.3)	(0.4)	(0.5)	(0.5)	(0.5)
TB Clinic	0.9	0.9	0.8	0.7	0.7
	(0.2)	(0.3)	(0.4)	(0.4)	(0.5)
Laboratory	0.9	0.9	0.4	0.4	0.1
	(0.3)	(0.3)	(0.5)	(0.5)	(0.3)
Number of clinics	18	55	51	60	421

Note: This table shows a comparison of clinic characteristics according to the year they began providing ART. Column (1) shows the clinics that began providing ART before 2005, and most of these facilities had ART before the national rollout. Column (5) shows the characteristics for clinics that have not begun providing ART as of the beginning of 2011. Standard deviations of the means are in parentheses.

Table A2 – Further robustness: Saving Behavior and Mortality Risk

	Has Savings			
	(1) Full smpl No controls	(2) Full smpl All controls	(3) Analysis smpl All controls	(4) Fixed smpl All controls
2010 × ART Proximity	0.09***	0.14**	0.10**	0.13**
	(0.03)	(0.06)	(0.05)	(0.06)
2008 × ART Proximity	0.09**	0.13***	0.07*	0.13***
	(0.04)	(0.05)	(0.04)	(0.05)
2010 × Clinic Proximity			−0.03	
			(0.03)	
2008 × Clinic Proximity			0.01	
			(0.03)	
Observations	4827	3483	3483	3345
R^2	0.08	0.12	0.10	0.13
	Total Savings ln(USD)			
2010 × ART Proximity	0.47***	0.84***	0.42*	0.84***
	(0.14)	(0.28)	(0.22)	(0.29)
2008 × ART Proximity	0.39**	0.70***	0.34	0.68***
	(0.17)	(0.26)	(0.21)	(0.26)
2010 × Clinic Proximity			−0.07	
			(0.13)	
2008 × Clinic Proximity			0.01	
			(0.14)	
Observations	4822	3481	3481	3345
R^2	0.12	0.15	0.14	0.16
	Mortality Risk Index			
2010 × ART Proximity	−0.24***	−0.22**	−0.05	−0.21**
	(0.08)	(0.09)	(0.08)	(0.10)
2008 × ART Proximity	−0.18*	0.07	0.15	0.11
	(0.09)	(0.10)	(0.09)	(0.10)
2010 × Clinic Proximity			0.00	
			(0.08)	
2008 × Clinic Proximity			−0.06	
			(0.05)	
Observations	4645	3416	3416	3345
R^2	0.04	0.25	0.22	0.25

* $p < .10$, ** $p < .05$, *** $p < .01$

Note: Standard errors (in parentheses) are clustered by village (113 clusters) and robust to heteroskedasticity. ART proximity is parameterized as the negative of log distance by road. All regressions include individual fixed effects and region-by-year dummies. The top panel shows regressions for the binary outcome of savings, the middle panel shows results for total savings, transformed using the inverse hyperbolic sine function, and lower panel shows results for the mortality risk index constructed using factor score. Columns 1 and 2 use the full sample of respondents. Column 3 is the analytical sample in the main text, restricted to individuals who were interviewed in all three years of the survey (2006, 2008, and 2010), however it is not more restrictive than the sample of individuals with complete information for the 3 years. Column 4 is restricted to individuals in who also have no missing observations in for our main outcomes, so the sample is fixed across regressions. Regressions with all controls include all spatial, demographic, visit month, and mean reversion controls interacted with year, as well as economic shocks and aid programs. Spatial controls include population density and region-specific proximity to clinic, market, major road and school (all interacted with $Year_t$). Column 3 also reports the coefficients on any clinic proximity, so proximity to any clinic is not region-specific in this specification. Demographic controls include pre-period wealth, roof material, if the respondent has a mobile phone, age, household size, gender, education, and marital status (all interacted with $Year_t$). Controls for economic shocks and other aid programs are described in detail in the text.

Table A3 – Further robustness: Investment Behavior

	Education ln(USD)			
	(1) Full smpl No controls	(2) Full smpl All controls	(3) Analysis smpl All controls	(4) Fixed smpl All controls
2010 × ART Proximity	0.36***	0.26	0.24**	0.31*
	(0.11)	(0.17)	(0.12)	(0.17)
2008 × ART Proximity	0.09	−0.03	0.04	−0.06
	(0.09)	(0.17)	(0.11)	(0.19)
2010 × Clinic Proximity			0.12	
			(0.09)	
2008 × Clinic Proximity			−0.02	
			(0.06)	
Observations	3396	2645	2553	2140
R^2	0.26	0.33	0.29	0.31
	Medical ln(USD)			
2010 × ART Proximity	0.13*	0.02	0.13	−0.03
	(0.08)	(0.09)	(0.08)	(0.09)
2008 × ART Proximity	0.04	0.09	0.06	0.01
	(0.06)	(0.13)	(0.08)	(0.15)
2010 × Clinic Proximity			−0.02	
			(0.05)	
2008 × Clinic Proximity			−0.01	
			(0.06)	
Observations	3475	2714	2606	2140
R^2	0.03	0.13	0.08	0.11
	Clothing ln(USD)			
2010 × ART Proximity	0.11	0.04	0.12	−0.01
	(0.13)	(0.15)	(0.13)	(0.16)
2008 × ART Proximity	−0.03	−0.03	−0.05	−0.17
	(0.11)	(0.21)	(0.14)	(0.24)
2010 × Clinic Proximity			0.09	
			(0.08)	
2008 × Clinic Proximity			−0.11*	
			(0.07)	
Observations	3474	2713	2605	2140
R^2	0.22	0.29	0.26	0.26

* $p < .10$, ** $p < .05$, *** $p < .01$

Note: Standard errors (in parentheses) are clustered by village (109 clusters) and robust to heteroskedasticity. ART proximity is parameterized as the negative of log distance by road. All regressions include individual fixed effects and region-by-year dummies. The top panel reports results for education expenditure outcome, the middle panel reports results for medical expenditure, and the bottom panel for clothing expenditure. All outcomes are transformed using the inverse hyperbolic sine transformation. The starting sample is all individuals who report biological children. Columns 1 and 2 use the full sample of respondents. Column 3 is the analytical sample in the main text, restricted to individuals who were interviewed in all three years of the survey (2006, 2008, and 2010), including data from 2004 where available. Column 4 is restricted to individuals in who also have no missing observations in for our main outcomes, so the sample is fixed across regressions. Regressions with all controls include all spatial, demographic, visit month, and mean reversion controls interacted with year, as well as economic shocks and aid programs. Spatial controls include population density and region-specific proximity to clinic, market, major road and school (all interacted with $Year_t$). Column 3 also reports the coefficients on any clinic proximity, so proximity to any clinic is not region-specific in this specification. Demographic controls include pre-period wealth, roof material, if the respondent has a mobile phone, age, household size, gender, education, and marital status (all interacted with $Year_t$). Controls for economic shocks and other aid programs are described in detail in the text.

Table A4 – ART Availability and Expectations – Individual Components of Mortality Risk Index

	Mortality Risk Index Components				Own Mortality Risk
	(1)	(2)	(3)	(4)	(5)
	Own	Healthy	HIV+	ART	P(HIV)>0 subsample
Panel A: 10-year horizon					
2010 × ART Proximity	−0.09	−0.41*	−0.42**	−0.43**	−1.08**
	(0.28)	(0.22)	(0.18)	(0.22)	(0.47)
2008 × ART Proximity	0.34	−0.07	−0.42**	−0.04	−0.76
	(0.26)	(0.22)	(0.20)	(0.23)	(0.46)
Mean dep. var	6.04	6.13	8.90	8.30	6.35
Observations	3698	3733	3734	3695	1139
R^2	0.08	0.08	0.08	0.12	0.12
Panel B: 5-year horizon					
2010 × ART Proximity	−0.42	−0.35*	−0.48	−0.31	−1.45***
	(0.26)	(0.20)	(0.30)	(0.24)	(0.50)
2008 × ART Proximity	0.00	0.02	−0.43	0.03	−1.22**
	(0.26)	(0.20)	(0.35)	(0.26)	(0.49)
Mean dep. var	4.19	4.15	7.04	6.25	4.49
Observations	3700	3737	3737	3698	1137
R^2	0.08	0.09	0.09	0.11	0.13
Panel C: 1-year horizon					
2010 × ART Proximity	−0.42*	−0.32*	−0.62***	0.04	−1.59***
	(0.23)	(0.16)	(0.24)	(0.22)	(0.50)
2008 × ART Proximity	−0.08	−0.04	−0.51*	0.24	−1.30***
	(0.23)	(0.15)	(0.29)	(0.22)	(0.49)
Mean dep. var	2.25	2.12	4.55	4.00	2.51
Observations	3705	3738	3739	3694	1138
Within R^2	0.08	0.09	0.08	0.09	0.13
Individual FEs	Y	Y	Y	Y	Y
Region × Year FEs	Y	Y	Y	Y	Y
Spatial controls	Y	Y	Y	Y	Y
Demo., shocks & aid	Y	Y	Y	Y	Y

* $p < .10$, ** $p < .05$, *** $p < .01$

Note: Standard errors (in parentheses) are clustered by village (113 clusters) and robust to heteroskedasticity. ART proximity is parameterized as the negative of log distance by road. All regressions include individual fixed effects, region-by-year dummies, and month of interview controls. The sample is restricted to individuals who were interviewed in all three years of the survey (2006, 2008, and 2010). Spatial controls include population density and proximity to clinic, market, major road and school (all interacted with region and $Post_t$). Demographic controls include pre-period wealth, roof material, if the respondent has a mobile phone, age, household size, gender, education, HIV status, and marital status (all interacted with $Post_t$). Controls for economic shocks and other aid programs are described in detail in the text. Each outcomes variable represents a probability, measured in increments of 1 between 0 and 10.

Table A5 – Heterogeneity and timing by pre-ART perceived HIV-risk

	(1) Mortality risk index	(2) Any savings
2010 × ART Proximity × $P(HIV)_{06} > 0$	−0.21	0.06
	(0.14)	(0.08)
2008 × ART Proximity × $P(HIV)_{06} > 0$	−0.28*	0.11
	(0.15)	(0.09)
2010 × ART Proximity	−0.21*	0.12**
	(0.11)	(0.05)
2008 × ART Proximity	0.03	0.08
	(0.12)	(0.06)
Observations	3623	3623
Within R^2	0.10	0.11

* $p < .10$, ** $p < .05$, *** $p < .01$

Note: This table shows heterogeneity of responses to ART proximity by whether the respondent reported any likelihood of HIV risk in 2006 (pre-ART). Standard errors (in parentheses) are clustered by village (113 clusters) and robust to heteroskedasticity. ART proximity is parameterized as the negative of log distance by road. All regressions include individual fixed effects, region-by-year dummies, and month of interview controls. The sample is restricted to individuals who were interviewed in all three years of the survey (2006, 2008, and 2010). Spatial controls include population density and proximity to clinic, market, major road and school (all interacted with region and $Post_t$). Demographic controls include pre-period wealth, roof material, if the respondent has a mobile phone, age, household size, gender, education, HIV status, and marital status (all interacted with $Post_t$). Controls for economic shocks and other aid programs are described in detail in the text. Mortality risk index is a factor score of a set of 16 variables describing own and hypothetical mortality risks.

Table A6 – Mortality risk for hypothetically healthy

	Hypothetical mortality risk for healthy person			Mortality risk for healthy HIV+ mort. risk partialled out		
	(1) 10-year horizon	(2) 5-year horizon	(3) 1-year horizon	(4) 10-year horizon	(5) 5-year horizon	(6) 1-year horizon
2010 × ART Proximity	−0.41*	−0.35*	−0.32*	−0.09	−0.05	−0.06
	(0.22)	(0.20)	(0.16)	(0.20)	(0.15)	(0.14)
2008 × ART Proximity	−0.07	0.02	−0.04	0.23	0.28	0.18
	(0.22)	(0.20)	(0.15)	(0.22)	(0.18)	(0.16)
Observations	3733	3737	3738	3731	3732	3732
R^2	0.08	0.09	0.09	0.04	0.03	0.03

* $p < .10$, ** $p < .05$, *** $p < .01$

Note: This table shows response to ART proximity for respondents' reporting of hypothetical mortality risk for a healthy individual over 3 time horizons. Columns 1-3 show the results for the 3 time horizons, whereas Columns 4-6 show the results after partialling out the effects of hypothetical HIV+ mortality risk . Standard errors (in parentheses) are clustered by village (113 clusters) and robust to heteroskedasticity. ART proximity is parameterized as the negative of log distance by road. All regressions include individual fixed effects, region-by-year dummies, and month of interview controls. The sample is restricted to individuals who were interviewed in all three years of the survey (2006, 2008, and 2010). Spatial controls include population density and proximity to clinic, market, major road and school (all interacted with region and $Post_t$). Demographic controls include pre-period wealth, roof material, if the respondent has a mobile phone, age, household size, gender, education, HIV status, and marital status (all interacted with $Post_t$). Controls for economic shocks and other aid programs are described in detail in the text. Mortality risk index is a factor score of a set of 16 variables describing own and hypothetical mortality risks.

Table A7 – Imputing savings: Modeling total cash savings in 2006

	Predictors of savings in 2006	
	Dependent variable: Total savings (ln)	
	β	(s.e.)
Demographics		
Male	0.26**	(0.12)
Education	0.01	(0.02)
Married	0.25	(0.16)
No. of children	−0.03*	(0.02)
Age	0.00	(0.02)
Age^2	−0.00	(0.00)
High discount rate	−0.23**	(0.11)
Income		
Earnings ln(USD)	0.14***	(0.04)
Land (hectares)	0.07*	(0.04)
No. goats	0.04**	(0.02)
No. pigs	0.01	(0.03)
No. chicken	0.01	(0.01)
No. cows	0.02	(0.01)
Has bed	0.33***	(0.12)
Has radio	0.19	(0.14)
Has bicycle	0.10	(0.12)
Has pit latrine	−0.00	(0.17)
Has paraffin lamp	0.16	(0.13)
Has metal roof	0.26	(0.16)
Expenditures		
Fertilizer ln(USD)	−0.06	(0.05)
Seed ln(USD)	0.19**	(0.10)
Hired labor ln(USD)	0.17***	(0.05)
Education ln(USD/child)	0.05	(0.08)
Medical ln(USD/child)	−0.22*	(0.13)
Clothing ln(USD/child)	0.20***	(0.06)
HIV/AIDS-related		
No. know died of AIDS	0.01	(0.01)
Worried about AIDS	0.11	(0.15)
HIV+	0.14	(0.29)
Observations	1096	
R^2	0.19	

* $p < .10$, ** $p < .05$, *** $p < .01$

Note: Standard errors (in parentheses) are clustered by village and robust to heteroskedasticity. This table presents the estimates used to impute savings. The estimation results are from modeling saving in 2006 as a function of the independent variables in 2006. The dependent variable is always total cash savings (transformed using inverse hyperbolic sine) in 2006, and the independent variables included to impute savings are listed in the the rows. Separate models were estimated for the subset of variables that were actually measured in 1998, 2001, and 2004 (results not shown). For example, to predict savings in 2004 when earnings were not reported, the regression used to obtain 2006 coefficients did not include earnings in 2006 since earnings were not reported in 2004.

Table A8 – Imputing versus measured savings in 2006-2010

	(1) Imputed savings ln(USD)	(2) Measured savings ln(USD)
2010 × ART Proximity	0.16	1.47***
	(0.12)	(0.27)
2008 × ART Proximity	0.13	1.21**
	(0.10)	(0.47)
Observations	2440	2433
R^2	0.31	0.15
Share of variance explained by imputed savings in:		
2008 (N=625)	0.22	
2010 (N=712)	0.29	

* $p < .10$, ** $p < .05$, *** $p < .01$
Note: Standard errors (in parentheses) are clustered by village and robust to heteroskedasticity. This table compares the difference in using imputed versus actual measures of savings to estimate the response to ART for years 2006, 2008, and 2010 (when actual measures of savings exist). The dependent variable is always total cash savings (transformed using inverse hyperbolic sine), either the imputed (Column 1) or actual (Column 2). The controls include individual FEs, region-by-year effects, and spatial controls (as in the rest of the paper). Imputed savings are calculated by modeling saving in 2006 as a function of the independent variables in 2006, savings the coefficients on the covariates, and then predicting or imputing savings based on the measured set of covariates for each survey wave (1998-2010). We do not have actual savings before the 2006 wave, but since we have actual savings from 2006-2010, we compare how well our imputed measure does at predicting actual savings in the second panel, which reports the share of variation in total savings that is explained by the imputed measure (ie, the R2 of regressing actual savings on the imputed measure).

Table A9 – Pre-trends in imputed savings, earnings, roof structure, AIDS worries and robustness to controlling for pre-trends

	Pre-ART trends: (Waves 1998, 2001, 2004, 2006)				Controlling for saving pre-trend: (Waves 2006, 2008, 2010)
	(1) Predicted savings ln(USD)	(2) Earnings ln(USD)	(3) Has metal roof	(4) Worried about AIDS	(5) Total Savings ln(USD)
Panel A: Pre-ART trends					
Year × ART Prox.	−0.03	−0.03	0.00	0.00	
	(0.03)	(0.03)	(0.01)	(0.01)	
Panel B: Controlling for pre-ART savings trend					
2010 × ART Proximity					0.86***
					(0.25)
2008 × ART Proximity					0.43
					(0.27)
Mean dep. var	1.01	4.26	0.12	0.50	1.55
Observations	3070	2736	4538	4309	3972
Within R^2	0.08	0.16	0.06	0.23	0.15
Individual FEs	Y	Y	Y	Y	Y
Region × Year FEs	Y	Y	Y	Y	Y
Spatial controls	Y	Y	Y	Y	Y

* $p < .10$, ** $p < .05$, *** $p < .01$

Note: Standard errors (in parentheses) are clustered by village and robust to heteroskedasticity. ART proximity is parameterized as the negative of log distance by road. All regressions include individual fixed effects, region-by-year dummies, month of interview controls, and spatial controls as described in the main text. Dependent variable in Column 1 uses imputed log-savings using variables described in Table **??** for years 1998-2004, but actual reported cash savings from 2006. Columns 2-4 estimate the pre-ART trends for earnings (transformed with the inverse hyperbolic sine, though earnings were not reported in 2004), roof structure, and worry about AIDS. Column 5 estimates the main savings results with the addition of a control for the pre-ART annual growth rate in savings interacted with year. As in the main text, column (5) sample is restricted to individuals who were interviewed in all three years of the survey (2006, 2008, and 2010).

Table A10 – Pre-ART (2006) characteristics of the MLSFH study population by attrition status

	Pre-ART (2006) Characteristics			Coefficient on ART Proximity:	
	(1) Analytic sample	(2) Attritors	(3) p-value diff (1)-(2)	(4) β	(5) p-value
Panel A: Socioeconomic Characteristics (2006)					
Age	36.85	33.74	0.00	−1.18	0.50
Household size	5.46	5.10	0.00	−0.19	0.55
Education (grades completed)	5.10	5.49	0.17	0.76	0.02**
Labor income ln(USD)	4.00	3.94	0.42	−0.44	0.06*
Land (hectares)	1.59	1.54	0.21	−0.16	0.48
High discount rate	0.67	0.65	0.64	−0.01	0.82
Wealth index (20 item)	0.11	0.15	0.72	0.61	0.13
Has metal roof	0.15	0.21	0.00	0.02	0.65
Has bicycle	0.58	0.60	0.21	0.05	0.53
Has radio	0.76	0.74	0.68	0.05	0.41
Has mobile phone	0.04	0.05	0.17	0.08	0.06*
Panel B: HIV, Health, & Risk Perceptions (2006)					
HIV Positive	0.04	0.10	0.00	0.05	0.23
Physical health score (PCS12)	52.50	52.18	1.00	−0.24	0.85
Mental health score (MCS12)	55.57	55.48	0.63	−1.64	0.10*
Know someone on ART	0.50	0.49	0.93	−0.19	0.04**
Worried about AIDS	0.27	0.30	0.48	0.03	0.59
Mortality risk (5 year; own)	3.87	4.02	0.06	0.62	0.14
Perceived likelihood of HIV (Likert)	0.36	0.43	0.17	0.05	0.72
Perceived HIV prevalence	0.28	0.30	0.11	0.02	0.30
Panel C: Savings & Expenditures on children (2006)					
Has savings	0.22	0.24	0.18	0.08	0.22
Total savings ln(USD)	0.92	1.08	0.06	0.36	0.29
Education ln(USD/child)	0.54	0.67	0.19	0.12	0.67
Clothing ln(USD/child)	1.06	1.08	0.54	0.25	0.30
Medical ln(USD/child)	0.28	0.38	0.00	−0.01	0.90
Panel E: Spatial Characteristics					
Distance to ART in 2006 (km)	26.42	26.65	0.71		
Distance to ART in 2008 (km)	9.10	8.87	0.11		
Distance to clinic (km)	6.11	5.72	0.02		
Distance to major market (km)	5.29	5.51	0.10		
Distance to major road (km)	4.98	4.32	0.00		
Distance to school (km)	1.66	1.67	0.91		
Population density (pers/km^2)	100.68	103.91	0.10		
Sample size	1379	716		709	
Joint test Panels A-C (p-value)				0.53	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
This table describes characteristics of respondents in 2006, before ART became available, for the sample of respondents in the analysis (column 1) and for those that were lost to followup (column 2). P-values in column 3 indicate significant differences between means in columns 1 and 2. Column 4 provides the coefficient from regressing each characteristic (in 2006) on ART proximity for the attritor sample (controlling for region and spatial characteristics), where each row represents one regression, and column 5 reports the p-value for the estimate.

Table A11 – Pre-ART characteristics of the MLSFH: attrition and migration followup

	Pre-ART (2006) Characteristics			Coefficient on ART Proximity:	
	(1)	(2)	(3)	(4)	(5)
	Analytic sample	Attritors found in 2010/2013	Attritors never located	Analytic sample+ attritors found β	p-val
Panel A: Socioeconomic Characteristics (2006)					
Age	36.85	31.70 (0.00)***	36.84 (0.99)	−0.59	0.51
Household size	5.46	5.00 (0.00)***	5.25 (0.18)	−0.16	0.45
Education (grades completed)	5.10	5.94 (0.00)***	4.80 (0.20)	0.33	0.18
Labor income ln(USD)	4.00	3.89 (0.27)	4.01 (0.94)	−0.16	0.39
Land (hectares)	1.59	1.57 (0.91)	1.49 (0.36)	−0.09	0.54
High discount rate	0.66	0.64 (0.63)	0.63 (0.38)	−0.06	0.14
Wealth index (20 item)	0.11	0.12 (0.92)	0.20 (0.53)	0.26	0.35
Has metal roof	0.15	0.20 (0.01)**	0.21 (0.02)**	0.10	0.03**
Has bicycle	0.58	0.55 (0.36)	0.67 (0.01)***	0.03	0.62
Has radio	0.76	0.72 (0.06)*	0.77 (0.92)	−0.02	0.66
Has mobile phone	0.04	0.07 (0.01)***	0.03 (0.86)	0.06	0.01**
Panel B: HIV, Health, & Risk Perceptions (2006)					
HIV Positive	0.04	0.09 (0.00)***	0.11 (0.00)***	0.01	0.41
Physical health score (PCS12)	52.50	52.62 (0.77)	51.55 (0.06)*	−0.77	0.22
Mental health score (MCS12)	55.57	56.03 (0.32)	54.70 (0.12)	−1.33	0.01**
Know someone on ART	0.50	0.49 (0.77)	0.48 (0.72)	−0.00	0.98
Worried about AIDS	0.27	0.32 (0.02)**	0.26 (0.72)	0.03	0.33
Mortality risk (5 year; own)	3.87	3.76 (0.45)	4.39 (0.00)***	0.25	0.14
Perceived likelihood of HIV (Likert)	0.36	0.44 (0.08)*	0.43 (0.13)	0.04	0.41
Perceived HIV prevalence	0.28	0.30 (0.19)	0.29 (0.47)	0.02	0.21
Panel C: Savings & Expenditures on children (2006)					
Has savings	0.22	0.23 (0.66)	0.26 (0.15)	−0.01	0.70
Total savings ln(USD)	0.92	1.01 (0.40)	1.18 (0.05)**	−0.08	0.67
Education ln(USD/child)	0.46	0.51 (0.31)	0.50 (0.46)	−0.00	0.98
Clothing ln(USD/child)	0.98	0.85 (0.04)**	0.90 (0.32)	−0.08	0.44
Medical ln(USD/child)	0.25	0.26 (0.82)	0.32 (0.06)*	−0.01	0.87
Panel E: Spatial Characteristics					
Distance to ART in 2006 (km)	26.42	26.98 (0.05)**	26.16 (0.45)		
Distance to ART in 2008 (km)	9.10	8.87 (0.25)	8.87 (0.32)		
Distance to clinic (km)	6.11	5.73 (0.03)**	5.71 (0.04)**		
Distance to major market (km)	5.29	5.21 (0.69)	5.97 (0.01)***		
Distance to major road (km)	4.98	4.36 (0.00)***	4.25 (0.00)***		
Distance to school (km)	1.66	1.70 (0.45)	1.64 (0.76)		
Population density (pers/km^2)	100.68	101.12 (0.89)	108.09 (0.05)**		
Sample size	1379	429	287	1808	
Joint test Panels A-C (p-value)				0.44	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
This table describes characteristics of respondents in 2006, before ART became available, for the sample of respondents in the main analysis (column 1), individuals who were temporarily attrited or migrated but were located in 2010 or 2013 (column 2), and for those that were entirely lost to followup (column 3). P-values in columns 2 and 3 indicate significant differences between means in those columns and column 1. Column 4 provides the coefficient from regressing each characteristic (in 2006) on ART proximity using the expanded sample to include attritors (ie, the sample from columns 1+2) who were found (controlling for region and spatial characteristics). Each row represents one regression, and column 5 reports the p-value for the estimate.

Table A12 – ART Availability and Saving Behavior – IPW Attrition Correction

	Any Savings				Savings (USD)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2010 × ART Proximity	0.07	0.13***	0.12**	0.10**	0.39**	0.86***	0.76***	0.66***
	(0.04)	(0.05)	(0.05)	(0.05)	(0.17)	(0.22)	(0.23)	(0.25)
2008 × ART Proximity	0.03	0.09*	0.12**	0.12**	0.17	0.64**	0.73***	0.69***
	(0.04)	(0.05)	(0.05)	(0.05)	(0.20)	(0.25)	(0.25)	(0.25)
Mean dep. var	0.32	0.32	0.32	0.32	1.55	1.55	1.54	1.54
Observations	3989	3989	3575	3473	3984	3984	3570	3470
Within R^2	0.07	0.09	0.10	0.10	0.11	0.12	0.13	0.14
Spatial controls	–	Y	Y	Y	–	Y	Y	Y
Demo. & mean rev. controls	–	–	Y	Y	–	–	Y	Y
Shocks & aid programs	–	–	Y	Y	–	–	Y	Y
Individual FEs	Y	Y	Y	Y	Y	Y	Y	Y
Region × Year FEs	Y	Y	Y	Y	Y	Y	Y	Y

* $p < .10$, ** $p < .05$, *** $p < .01$

Note: Standard errors (in parentheses) are clustered by village (113 clusters) and robust to heteroskedasticity. ART proximity is parameterized as the negative of log distance by road. All regressions include individual fixed effects, interview month controls, and region-by-year dummies. The sample is restricted to individuals who were interviewed in all three years of the survey (2006, 2008, and 2010). Spatial controls include population density and region-specific proximity to clinic, market, major road and school (all interacted with $Post_t$). Demographic controls include pre-period wealth, roof material, if the respondent has a mobile phone, age, household size, gender, education, HIV status, and marital status (all interacted with $Post_t$). Mean reversion controls include pre-ART mortality risk and perceived HIV risk, interacted with $Year_t$. Controls for economic shocks and other aid programs are described in detail in the text.

Table A13 – ART Availability and Expenditures on Children – IPW Attrition Correction

	Education				Medical				Clothing			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
2010 × ART Proximity	0.48***	0.49***	0.40**	0.50***	0.17*	0.23**	0.09	0.19*	0.13	0.11	0.12	0.22
	(0.13)	(0.16)	(0.16)	(0.19)	(0.09)	(0.11)	(0.10)	(0.11)	(0.13)	(0.15)	(0.15)	(0.20)
2008 × ART Proximity	0.15	0.17	0.03	0.13	0.13	0.19	0.11	0.21**	0.08	0.05	−0.09	0.02
	(0.12)	(0.17)	(0.18)	(0.20)	(0.11)	(0.14)	(0.12)	(0.11)	(0.14)	(0.19)	(0.18)	(0.19)
2006 × ART Proximity				0.13				0.13*				0.14
				(0.14)				(0.08)				(0.18)
Mean dep. var	1.02	1.02	1.02	1.02	0.36	0.36	0.37	0.37	1.44	1.44	1.45	1.45
Observations	2833	2833	2543	2543	2890	2890	2596	2596	2889	2889	2595	2595
Within R^2	0.26	0.27	0.29	0.29	0.04	0.05	0.09	0.09	0.23	0.24	0.26	0.26
Spatial controls	–	Y	Y	Y	–	Y	Y	Y	–	Y	Y	Y
Demo. & mean rev. controls	–	–	Y	Y	–	–	Y	Y	–	–	Y	Y
Shocks & aid programs	–	–	Y	Y	–	–	Y	Y	–	–	Y	Y
Household FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Region × Year FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

* $p < .10$, ** $p < .05$, *** $p < .01$

Note: Standard errors (in parentheses) are clustered by village (109 clusters) and robust to heteroskedasticity. ART proximity is parameterized as the negative of log distance by road. All regressions include individual fixed effects, region-by-year dummies, and month of interview controls. The sample is restricted to respondents with school-age children and regressions are weighted by inverse of number of household respondents. All regressions use data from 2004. Spatial controls include population density and region-specific proximity to clinic, market, major road and school (all interacted with $Post_t$). Demographic controls include pre-period wealth, roof material, if the respondent has a mobile phone, age, household size, gender, education, HIV status, and marital status (all interacted with $Post_t$). Mean reversion controls include pre-ART mortality risk and perceived HIV risk, interacted with $Year_t$. Controls for economic shocks and other aid programs are described in detail in the text.

Table A14 – Attrition Followup: Including attritors who were found in later waves or migration followup in 2013

	Any Savings			Savings (USD)			Mortality Risk Index		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
2010 × ART Proximity	0.14***	0.11***	0.12***	0.91***	0.70***	0.71***	−0.18**	−0.23***	−0.23***
	(0.04)	(0.04)	(0.04)	(0.21)	(0.23)	(0.24)	(0.09)	(0.08)	(0.08)
2008 × ART Proximity	0.11**	0.10**	0.10**	0.69***	0.64**	0.64**	−0.00	−0.05	−0.06
	(0.05)	(0.05)	(0.05)	(0.25)	(0.27)	(0.26)	(0.11)	(0.09)	(0.09)
Observations	3989	4836	4835	3984	4831	4830	3871	4643	4642
Within R^2	0.09	0.10	0.09	0.13	0.19	0.18	0.07	0.07	0.07
Sample	Baseline	+Attritors	+Attritors	Baseline	+Attritors	+Attritors	Baseline	+Attritors	+Attritors
IPW	–	–	Y	–	–	Y	–	–	Y
Spatial controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Individual FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y
Region × Year FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y

* $p < .10$, ** $p < .05$, *** $p < .01$

Note: Standard errors (in parentheses) are clustered by village (113 clusters) and robust to heteroskedasticity. ART proximity is parameterized as the negative of log distance by road. All regressions include individual fixed effects, interview month controls, and region-by-year dummies. The baseline sample is restricted to individuals who were interviewed in all three years of the survey (2006, 2008, and 2010). The baseline + attritors sample includes additionally the 429 individuals who attrited after 2006 but were in the survey in 2010 or were located in the migration followup in 2013 (essentially imputing 2010 values of the outcomes from the 2013 questionnaire). Spatial controls include population density and region-specific proximity to clinic, market, major road and school (all interacted with region and $Post_t$).

Table A15 – Pre-ART (2006) characteristics of the MLSFH study population (Winsorized)

			Coefficient on ART Proximity:			
	(1) Mean	(2) St.dev.	(3) β	(4) p-value	(5) β	(6) p-value
Panel A: Socioeconomic Characteristics (2006)						
Age	36.85	13.1	−0.46	0.46	0.04	0.97
Household size	5.46	2.3	−0.17	0.19	−0.12	0.61
Education (grades completed)	5.10	3.5	0.24	0.16	0.25	0.36
Labor income (USD)	80.58	215.8	21.42	0.08*	11.83	0.61
Land (hectares)	1.59	1.5	−0.05	0.62	0.01	0.97
High discount rate	0.67	0.5	−0.02	0.56	−0.04	0.35
Wealth index (20 item)	0.11	2.0	0.20	0.13	0.15	0.57
Has metal roof	0.15	0.4	0.07	0.02**	0.11	0.04**
Has bicycle	0.58	0.5	0.04	0.27	0.00	0.99
Has radio	0.76	0.4	0.02	0.55	−0.01	0.76
Has mobile phone	0.04	0.2	0.02	0.25	0.04	0.02**
Panel B: HIV, Health, & Risk Perceptions (2006)						
HIV Positive	0.04	0.2	−0.01	0.30	0.01	0.75
Physical health score (PCS12)	52.50	7.2	−0.50	0.18	−0.84	0.29
Mental health score (MCS12)	55.57	8.0	−1.03	0.02**	−0.95	0.16
Know someone on ART	0.50	0.5	0.04	0.34	0.02	0.79
Worried about AIDS	0.27	0.4	0.03	0.12	0.02	0.54
Mortality risk (5 year; own)	3.87	2.3	0.46	0.01***	0.20	0.28
Perceived likelihood of HIV (Likert)	0.36	0.7	0.11	0.00***	0.04	0.44
Perceived HIV prevalence	0.28	0.2	0.01	0.18	0.02	0.17
Panel C: Savings & Expenditures on children (2006)						
Has savings	0.22	0.4	−0.03	0.27	−0.04	0.32
Total savings (USD)	13.21	40.6	2.61	0.26	−2.32	0.56
Education (USD/child)	1.17	3.0	−0.02	0.91	0.02	0.95
Clothing (USD/child)	2.06	2.8	0.16	0.46	0.13	0.69
Medical (USD/child)	0.33	0.6	0.03	0.35	0.05	0.36
Panel D: Child outcomes & characteristics (2006)						
Child age	10.0	2.8	0.34	0.15	0.38	0.48
Grades completed	2.7	2.0	0.27	0.21	0.08	0.88
Panel E: Spatial Characteristics						
Distance to ART in 2006 (km)	26.4	4.9	4.02	0.00***		
Distance to ART in 2008 (km)	9.1	3.6	−6.84	0.00***		
Distance to clinic (km)	6.1	3.1	−3.48	0.00***		
Distance to major market (km)	5.3	3.8	−3.72	0.00***		
Distance to major road (km)	5.0	3.5	−1.78	0.00***		
Distance to school (km)	1.7	1.0	−0.09	0.58		
Population Density (pers/km^2)	101	56	−16.0	0.04**		
Joint test Panels A-C (p-value)	–	–	0.04		0.34	

$N = 1379$. For child sample (Panel D), $N = 525$. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: This table describes characteristics of respondents and their children in 2006, before ART became available. Columns (3) and (5) report the coefficient on ART proximity (parameterized as the negative log distance) from regressing each variable in 2006 on ART proximity, with respective p-values in columns (4) and (6). Column (3) only controls for region dummies, while column (5) controls for spatial characteristics listed in Panel E (described in detail in the text). The sample of survey respondents is restricted to those who were interviewed in all three years for the main analysis (2006, 2008, and 2010). Panel D describes characteristics of the respondents' children and is restricted to children who were reported in the household roster for all three years.

Table A16 – Pre-ART Trends: Changes between 2004–2006 (Winsorized)

			Coefficient on ART Proximity:			
	(1) Mean	(2) St.dev.	(3) β	(4) p-value	(5) β	(6) p-value
Panel A: Demographic and Economic Characteristics						
Household size	−0.42	2.4	−0.34	0.01**	−0.46	0.09*
Land (hectares)	0.41	1.6	0.09	0.58	0.20	0.33
High discount rate	0.20	0.6	−0.02	0.66	0.05	0.36
Wealth index (20 item)	0.04	1.2	−0.07	0.31	−0.09	0.50
Has metal roof	0.01	0.3	0.01	0.54	0.03	0.14
Has bicycle	0.04	0.5	−0.02	0.43	−0.03	0.47
Has radio	0.02	0.5	−0.03	0.29	−0.05	0.35
Has mobile phone	0.03	0.2	0.02	0.14	0.04	0.03**
Panel B: HIV, Health, and Risk Perceptions						
HIV Positive	0.01	0.1	0.00	0.59	−0.01	0.05*
Know someone on ART	−0.18	0.4	0.02	0.58	−0.02	0.75
Worried about AIDS	−0.27	0.5	0.06	0.04**	0.07	0.27
Perceived likelihood of HIV (Likert)	−0.25	1.1	0.10	0.16	−0.04	0.72
Perceived HIV prevalence	−0.12	0.3	0.01	0.61	−0.04	0.24
Panel C: Expenditures on Children						
Education (USD/child)	0.33	1.9	0.28	0.25	0.04	0.93
Clothing (USD/child)	0.34	3.5	−0.32	0.44	0.31	0.56
Medical (USD/child)	−0.02	0.9	0.07	0.30	0.17	0.16
Joint test Panels A-C (p-value)	–	–	0.21		0.43	

$N = 1354$ * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Note: This table shows the mean changes between 2004 and 2006 (i.e., the period before ART came online in the MLSFH study regions) in available outcomes and characteristics of the sample. Columns (3) and (5) report the coefficient on ART proximity (parameterized as the negative log distance) from regressing each variable in 2006 on ART proximity, with respective p-values in columns (4) and (6). Column (3) only controls for region dummies, while column (5) controls for spatial characteristics.

Table A17 – ART Availability and Saving Behavior (Winsorized)

	Any Savings				ln(Savings)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2010 × ART Proximity	0.10***	0.14***	0.13***	0.12**	30.82***	39.40**	32.98**	20.20
	(0.04)	(0.04)	(0.05)	(0.05)	(10.73)	(17.21)	(16.21)	(16.86)
2008 × ART Proximity	0.07*	0.11**	0.14***	0.13***	13.03*	20.66	17.32	8.85
	(0.04)	(0.05)	(0.05)	(0.05)	(7.39)	(13.17)	(11.53)	(12.28)
Mean dep. var	0.32	0.32	0.32	0.32	50.33	50.33	48.59	49.10
Observations	3989	3989	3575	3473	3984	3984	3570	3470
Within R^2	0.08	0.09	0.10	0.11	0.13	0.13	0.19	0.18
Spatial controls	–	Y	Y	Y	–	Y	Y	Y
Demo. controls	–	–	Y	Y	–	–	Y	Y
Shocks & aid programs	–	–	–	Y	–	–	–	Y
Household FEs	Y	Y	Y	Y	Y	Y	Y	Y
Region × Year FEs	Y	Y	Y	Y	Y	Y	Y	Y

* $p < .10$, ** $p < .05$, *** $p < .01$

Note: Standard errors (in parentheses) are clustered by village (113 clusters) and robust to heteroskedasticity. ART proximity is parameterized as the negative of log distance by road. All regressions include individual fixed effects and region-by-year dummies. The sample is restricted to individuals who were interviewed in all three years of the survey (2006, 2008, and 2010). Spatial controls include population density and proximity to clinic, market, major road and school (all interacted with $Post_t$). Demographic controls include pre-period wealth, roof material, if the respondent has a mobile phone, age, household size, gender, education, and marital status (all interacted with $Post_t$). Controls for economic shocks and other aid programs are described in detail in the text.

Table A18 – ART Availability and Expenditures on Children (Winsorized)

	Education				Medical				Clothing			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
2010 × ART Proximity	2.46**	2.87**	1.94	1.64	0.26*	0.20	0.06	0.21	0.07	0.23	0.30	0.41
	(1.03)	(1.33)	(1.19)	(1.28)	(0.14)	(0.16)	(0.17)	(0.18)	(0.76)	(0.88)	(0.82)	(0.90)
2008 × ART Proximity	0.53	1.11	−0.04	−0.34	0.13	0.08	−0.01	0.15	0.15	0.30	−0.17	−0.05
	(0.66)	(1.21)	(1.25)	(1.31)	(0.11)	(0.15)	(0.16)	(0.15)	(0.75)	(1.03)	(1.05)	(1.03)
2006 × ART Proximity				−0.37				0.20*				0.15
				(0.70)				(0.10)				(0.69)
Mean dep. var	3.49	3.49	3.47	3.47	0.48	0.48	0.49	0.49	4.25	4.25	4.32	4.32
Observations	2833	2833	2543	2543	2890	2890	2596	2596	2889	2889	2595	2595
Within R^2	0.13	0.14	0.19	0.19	0.04	0.05	0.08	0.08	0.16	0.17	0.21	0.21
Spatial controls	–	Y	Y	Y	–	Y	Y	Y	–	Y	Y	Y
Demo. controls	–	–	Y	Y	–	–	Y	Y	–	–	Y	Y
Shocks & aid programs	–	–	Y	Y	–	–	Y	Y	–	–	Y	Y
Household FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Region × Year FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

* $p < .10$, ** $p < .05$, *** $p < .01$
Note: Standard errors (in parentheses) are clustered by village (109 clusters) and robust to heteroskedasticity. ART proximity is parameterized as the negative of log distance by road. All regressions include individual fixed effects, region-by-year dummies, and month of interview controls. The sample is restricted to respondents with school-age children and regressions are weighted by inverse of number of household respondents. All regressions use data from 2004. Spatial controls include population density and proximity to clinic, market, major road and school (all interacted with $Post_t$). Demographic controls include pre-period wealth, roof material, if the respondent has a mobile phone, age, household size, gender, education, and marital status (all interacted with $Post_t$). Controls for economic shocks and other aid programs are described in detail in the text.

Table A19 – ART Availability, Income, and Other Expenditures (Winsorized)

	Expenditures				Farm Inputs				Income and Wealth		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Clothing (Own)	Medical (Own)	Medical (Others)	Funeral	Seed	Farm Equipt	Fertilizer	Hired Labor	Total Expend. ln(USD)	Labor Earnings ln(USD)	Wealth Index
2010 × ART Proximity	−1.66	0.11	0.24	−0.02	−0.11	−0.08	−5.57	−2.39	−0.34**	−0.05	−0.21
	(1.91)	(0.76)	(0.27)	(0.28)	(0.39)	(0.70)	(5.54)	(2.62)	(0.15)	(0.18)	(0.14)
2008 × ART Proximity	−1.60	−0.51	−0.09	0.03	−0.11	0.04	1.32	0.26	−0.20	−0.31	−0.23*
	(1.94)	(0.43)	(0.26)	(0.33)	(0.23)	(0.40)	(3.39)	(2.58)	(0.17)	(0.20)	(0.13)
2006 × ART Proximity	−0.12	0.15		−0.09	0.15	0.25*	3.32**	1.50			−0.16**
	(0.75)	(0.26)		(0.14)	(0.13)	(0.15)	(1.33)	(1.05)			(0.08)
Mean dep. var	10.92	1.95	0.66	1.16	0.77	0.99	8.72	6.11	3.29	4.92	0.13
Observations	5310	5307	3971	5267	5270	5268	5263	5272	3932	3994	4907
Within R^2	0.17	0.09	0.03	0.07	0.08	0.07	0.11	0.09	0.13	0.35	0.03
Household FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Region × Year FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Spatial controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

* $p < .10$, ** $p < .05$, *** $p < .01$

Note: Standard errors (in parentheses) are clustered by village (113 clusters) and robust to heteroskedasticity. Household level regressions (3-11) are inverse weighted by the number of respondents from that household. ART proximity is parameterized as the negative of log distance by road. All regressions include individual fixed effects, region-by-year dummies, and month of interview controls. The sample is restricted to individuals who were interviewed in all three years of the survey (2006, 2008, and 2010). Estimates for columns 1,2, 4-8, and 11 also use data from 2004. All monetary values are expressed in USD. Because of frequent zeros, the estimates in columns 1-8 use levels of total spending as the dependent variable, winsorized at 98th percentile. Total expenditures, Column 9, is the sum of expenditures in columns 1-8. Total expenditures and earnings, columns 9 and 10, are transformed using the inverse hyperbolic sine. The dependent variable in column 11 is a total wealth index (roughly mean 0, standard deviated 2) calculated using PCA incorporating all assets, housing and roofing structure, and animals.

Table A20 – ART Availability and Investment – HIV Mortality, Caretaking, and Orphans (Winsorized)

	Saving Behavior		Expenditures on Children		
	(1) Any Savings	(2) Savings (USD)	(3) Education	(4) Medical	(5) Clothing
Panel A: Aggregate Effect of ART					
2010 × ART Proximity	0.14***	39.40**	2.87**	0.20	0.23
	(0.04)	(17.21)	(1.33)	(0.16)	(0.88)
2008 × ART Proximity	0.11**	20.66	1.11	0.08	0.30
	(0.05)	(13.17)	(1.21)	(0.15)	(1.03)
Panel B: HIV-negative					
2010 × ART Proximity	0.18***	40.10**	1.85	0.12	0.28
	(0.04)	(17.66)	(1.33)	(0.17)	(0.89)
2008 × ART Proximity	0.13***	21.10	0.95	−0.04	−0.10
	(0.05)	(12.97)	(1.31)	(0.14)	(0.98)
Panel C: No family illness					
2010 × ART Proximity	0.14***	37.69*	2.84**	0.16	0.36
	(0.04)	(19.36)	(1.44)	(0.17)	(0.90)
2008 × ART Proximity	0.10**	18.67	1.28	0.01	0.36
	(0.05)	(13.20)	(1.31)	(0.15)	(1.07)
Panel D: No recent death					
2010 × ART Proximity	0.17***	25.56	2.37	0.34	0.13
	(0.05)	(17.33)	(1.75)	(0.22)	(1.09)
2008 × ART Proximity	0.12**	17.75	0.63	−0.11	−0.23
	(0.05)	(14.76)	(1.24)	(0.14)	(1.09)
Panel E: No orphans					
2010 × ART Proximity	0.16***	50.40**	2.65	0.35**	0.60
	(0.05)	(21.80)	(1.93)	(0.17)	(1.15)
2008 × ART Proximity	0.12**	14.05	0.71	0.25	0.46
	(0.06)	(14.10)	(1.27)	(0.20)	(1.37)
Panel F: No HIV/illness/death or orphans					
2010 × ART Proximity	0.20***	52.28***	1.84	0.37	0.37
	(0.06)	(20.21)	(1.69)	(0.28)	(1.58)
2008 × ART Proximity	0.10	16.86	1.29	0.08	0.29
	(0.07)	(15.83)	(1.55)	(0.16)	(1.39)
Household FEs	Y	Y	Y	Y	Y
Region × Year FEs	Y	Y	Y	Y	Y
Spatial controls	Y	Y	Y	Y	Y

* $p < .10$, ** $p < .05$, *** $p < .01$

Note: Standard errors (in parentheses) are clustered by village (columns 1 and 2 have 113 clusters, columns 3-5 have 109 clusters) and robust to heteroskedasticity. ART proximity is parameterized as the negative of log distance by road. All regressions include individual fixed effects and region-by-year dummies. The sample is restricted to individuals who were interviewed in all three years of the survey (2006, 2008, and 2010). Regressions reported in columns 3-5 are restricted to respondents with school-age children and weighted by inverse of number of household respondents. Columns 3-5 also use data from 2004. All regressions exclude HIV-positive respondents.

References

Aghion, Philippe, Diego Comin, Peter Howitt, and Isabel Tecu, "When Does Domestic Savings Matter for Economic Growth?," *IMF Economic Review*, jan 2016.

Anglewicz, Philip, Mark Vanlandingham, Lucinda Manda-Taylor, and Hans-Peter Kohler, "Cohort Profile : The Migration and Health in Malawi (MHM) Study," *PSC Working Paper Series*, 2015, (62).

Baranov, Victoria, Daniel Bennett, and Hans-Peter Kohler, "The indirect impact of antiretroviral therapy: Mortality risk, mental health, and HIV-negative labor supply," *Journal of Health Economics*, dec 2015, *44*, 195–211.

Barden-O'Fallon, Janine L., Joseph Degraft-Johnson, Thomas Bisika, Sara Sulzbach, Aimee Benson, and Amy O. Tsui, "Factors associated with HIV/AIDS knowledge and risk perception in rural Malawi," *AIDS and Behavior*, 2004, *8* (2), 131–140.

Blacker, John and Basia Zaba, "HIV prevalence and the lifetime risk of dying of AIDS," *Health Transition Review*, 1997, *7* (S2), 45–62.

Castro, Ruben, Jere R Behrman, and Hans-Peter Kohler, "Perception of HIV risk and the quantity and quality of children: the case of rural Malawi," *Journal of Population Economics*, jan 2015, *28* (1), 113–132.

Delavande, Adeline and Hans-Peter Kohler, "Subjective Expectations in the Context of HIV/AIDS in the Malawi," *Demographic Research*, jun 2009, *20* (31), 817–874.

Derksen, Laura and Adamson Muula, "Love in the Time of HIV : Theory and Evidence on Social Stigma and Health Seeking Behavior," 2016.

Hall, Peter, "Effect of Bias Estimation on Coverage Accuracy of Bootstrap Confidence Intervals for a Probability Density," *The Annals of Statistics*, 1992, *20* (2), pp. 675–694.

Kalipeni, Ezekiel and Jayati Ghosh, "Concern and practice among men about HIV/AIDS in low socioeconomic income areas of Lilongwe, Malawi," *Social Science & Medicine*, mar 2007, *64* (5), 1116–1127.

Kerwin, Jason T, "Scared Straight or Scared to Death? The Effect of Risk Beliefs on Risky Behaviors," *Unpublished manuscript*, 2016.

Kohler, H.-P., Susan C. Watkins, Jere R Behrman, Philip Anglewicz, Iliana V Kohler, R. L. Thornton, J. Mkandawire, H. Honde, A. Hawara, B. Chilima, C. Bandawe, V. Mwapasa, P. Fleming, and L. Kalilani-Phiri, "Cohort Profile: The Malawi Longitudinal Study of Families and Health (MLSFH)," *International Journal of Epidemiology*, apr 2015, *44* (2), 394–404.

Santos Silva, J. M C and Silvana Tenreyro, "Poisson: Some convergence issues," *Stata Journal*, 2011, *11* (2), 207–212.

Santos Silva, J.M.C. and Silvana Tenreyro, "On the existence of the maximum likelihood estimates in Poisson regression," *Economics Letters*, 2010, *107* (2), 310–312.

Tatem, Andrew and Catherine Linard, "Population mapping of poor countries.," *Nature*, jun 2011, *474* (7349), 36.

Tatem, Andrew J., Abdisalan M. Noor, Craig von Hagen, Antonio Di Gregorio, and Simon I. Hay, "High Resolution Population Maps for Low Income Nations: Combining Land Cover and Census in East Africa," *PLoS ONE*, dec 2007, *2* (12), e1298.

Thornton, Rebecca L, "The Demand for, and Impact of, Learning HIV Status.," *American Economic Review*, dec 2008, *98* (5), 1829–1863.

United Nations Population Division, *World Population Prospects, the 2010 Revision: Standard (Median) Forecasts*, United Nations, Department of Economic and Social Affairs, Population Division, 2010.

Wooldridge, Jeffrey M., "Inverse probability weighted estimation for general missing data problems," *Journal of Econometrics*, 2007, *141* (2), 1281–1301.

WorldPop Project, "Afripop Population Density Database," *http://www.afripop.org*, 2010.

Zhang, Jie Junsen and Jie Junsen Zhang, "The Effect of Life Expectancy on Fertility, Saving, Schooling, and Economic Growth: Theory and Evidence," *Scandinavian Journal of Economics*, mar 2005, *107* (1), 45–66.