Chronic Poverty
Research Centre

Toolkit Note

# Testing and adjusting for attrition in household panel data

Bob Baulch (CPRC)
Agnes Quisumbing (IFPRI)

## Abstract

This note describes the use of a simple procedure to correct for attrition due to observables in household panel survey: inverse probability weights. The procedure involves estimating two probit regressions, one with and one without variables that are significantly associated with attrition, and using the ratio of predicted probabilities from these regressions to reweight the observations. The procedure is illustrated in Stata using data from part of the CPRC-DATA-IFPRI panel in rural Bangladesh.

## Introduction

Attrition has been described as 'the panel researcher's nightmare' (Winkels and Withers, 2000). This is because if the members who drop out of a panel differ systematically from those who stay in it, then the dataset of continuing members is no longer representative of the original population. So results based on data in which only continuing panel members are included may be seriously affected by attrition bias. Fortunately, a number of studies suggest there is a simple method of adjusting for sample attrition (at least when it is based on observable characteristics) known as inverse probability weights (Fitzgerald *et al.*, 1988; Wooldridge, 2002). This Toolkit note aims to provide a simple introduction and illustration of how to test and adjust for attrition bias in panel data using the statistical software Stata.

To fix ideas consider a household panel consisting of i….N households who have been surveyed in two different years (t=1, 2). Denoting the outcome of interest for household i in the second year by $y_{i2}$, household variables in the first year by $x_{i1,}$ and additional instrumental variables that only affect attrition by $z_{i1}$, we may write:

(1) $y_{i2} = x_{i1}\beta + \varepsilon_i$ $y_{i2}$ observed if A*>1

(2) $A* = x_{i1}\gamma \ + z_{i1}\delta + v_i$

This set-up resembles a standard one-period selection model except that the outcome variable is measured in the second period. In practice, however, the probability of attrition, A*, is not observed and is replaced by an attrition dummy, A, which takes the value 0 when both $y_{i1}$ and $y_{i2}$ are observed, and one when $y_{i2}$ is not observed. So one possible solution to sample attrition is to estimate a selection model, which relies on identifying a set of instrumental variables, $z_i$ , which are correlated with attrition but not with $\varepsilon_i$ (Heckman, 1979).[1] This is often referred to the case of selection on unobservables. However, it is often difficult to identify suitable instrumental variables for selection models. A second solution to sample attrition is to estimate inverse probability weights, which relies on an auxiliary variable(s) which can be related to both attrition and the outcome variable (Fitzgerald et al, 1998). This is the case of selection on observables, and requires a much weaker condition for the z variables: that $\varepsilon_i$ and $v_i$ are uncorrelated. To estimate such inverse probability weights, equation (2) is respecified as a probit:

(3) $$A = x_{i1}\gamma + a_{i1}\delta + v_i$$

where A=0 for households who remain in the sample and A=1 for attritors, and $a_{i1}$ are the auxiliary variables in the first period. Next, a restricted version of equation is re-estimated without the auxiliary variables:

(4) $$A = x_{i1}\gamma + \varphi_i$$

The ratio of the predicted values from equation (4) and equation (3) give the inverse probability weights:

(5) $$W_i = \frac{p^r}{p^u}$$

The intuition behind this procedure is that it gives more weight to households who have similar initial characteristics to households that subsequently attrit than to households with characteristics that make them more likely to remain in the panel.

The question that then arises is what variables are suitable for inclusion in $z_i$ or $a_i$? Clearly these variables must be observed for both panel households and attritors, and be correlated with the probability of attrition. In selection models, lagged values of the outcome variable are often used as instrumental variables, but this typically requires that at least three waves of panel data are available. Measures of the quality of the interview are often included among the z variables (Maluccio, 2004) as these seem likely to be related to the probability of attrition but are not necessarily to the outcome variable. The auxiliary variables used to calculate inverse probability weights can also include household demographic variables, community level variables, shocks or treatment variables (for example, whether a household

[1] Note that $\varepsilon_i$ and $v_i$ are not uncorrelated in this case.

receives a transfer payment). Attrition is often found to be related to the age of the household head, or the demographic composition of the household and also to shocks or treatment variables. As these variables are usually correlated with the outcomes, demographic, community and shock variables cannot be used in the selection model in equations (1) and (2) –which requires z to be instrumental variables which are correlated with attrition but not the outcome variable. However, they can be used in the $a_i$ that are used when calculating inverse probability weights. In our application below, we include a variable for community level shocks and households' technology adoption status in year 1 as well as the household demographic and thana (sub-district) dummies.

## Testing whether attrition is random

Prior to calculating inverse probability weights, it is first essential to test whether attrition in a panel data model is random. There may be situations in which attrition is entirely random and, in this fortunate situation, it is not necessary to do anything further.

A number of tests have been proposed for whether attrition in a panel is random, including attrition probits (Fitzgerald et al, 1998) and pooling tests, in which the equality of coefficients from the baseline sample with and without attritors are equal (Becketti, Gould, Lillard and Welch, 1988). We implement both of these tests in this note.[2]

The simplest test for whether attrition is random is to estimate a probit in which the dependent variables takes the value one for households which drop out of the sample after the first wave (attrit) and zero otherwise. Explanatory variables are baseline values for all variables that are believed to affect the outcome variable of interest plus any available variables which characterise the interview process. It is usual to include lagged values of the outcome variable in such attrition probits. As pointed out by Outes-Leon and Dercon (2008), it is also useful to examine the pseudo R-squared from attrition probits, as they can be interpreted as the proportion of attrition that is non-random.

Another commonly used test for whether attrition is random is the pooling test due to Becketti, Gould, Lillard and Welch (1988). The Becketti, Gould, Lillard and Welch (hereafter BGLW) test involves regressing an outcome variable from the first wave of a survey on household and community variables, an attrition dummy, and the attrition dummy interacted with the other explanatory variables. An F-test of the joint significance of the attrition dummy and the interaction variables is then conducted to determine whether the coefficients from the explanatory variables differ between households who are retain or attrit from the panel.

[2] Other tests for attrition include parametric selectivity models (Falaris, 2003) and examining the significance of lagged dependent variables.

It is important to understand that these tests are model specific and needs to be repeated for each outcome variable of interest. Thus in our application to data from rural households in Bangladesh, we calculate separate tests for expenditures and assets.

## Application

To illustrate the calculation of inverse probability weights we use the agricultural technologies portion of the CPRC-DATA-IFPRI panel from rural Bangladesh. This panel spans the ten year period from late 1996 to late 2006/early 2007, and contains just over 1,300 households located in four of Bangladesh's 64 districts: Manikganj, Kishoreganj, Jessore, and Mymensingh. The survey was clustered at the village level, and there are 47 villages included in it (see http://www.ifpri.org/dataset/chronic-poverty-and-long-term-impact-study-bangladesh for further details of this dataset, which is publicly accessible).

The Stata dataset Bangladesh_example.dta contains observations for 965 households in 1996 (the 'baseline'), 47 of whom drop out of the sample between 1996 and 2006. Thus attrition from the panel at the household level is just under 5%. There are also 10 households for whom we do not have expenditure data in 1996, who are included among the attritors in the expenditure model below.[3] In addition to the usual variables on the household's demographic characteristics, age and education of the household head, asset ownership and location (thana) variables, we have detailed information on four variables that may be correlated with attrition. These are: (i) lagged values of the dependent variable (per capita expenditures or the value of assets owned by the household); (ii) the percentage of households in the village experiencing a flood between 1996 and 2006; (iii) the village level attrition rate during the four rounds of the survey conducted in 1996, which is taken as a indicator of interview quality;[4] and, (iv) the adoption status of household with respect to the agricultural technologies (introduced vegetables, individual and group fishponds) that were the focus of the original study. All these variables are observed for both attritors and households that remain in the sample. Note that with the exception of the village attrition rate, these variables could not be included in a selection type model, as they are correlated with both attrition and the outcome variable

Each of these variables are included in the attrition probit for expenditures produced by the following Stata command and reported in Table 1 below:

[3] Note that households who split (sub-divided) between 1996 and 2007 have been excluded from the dataset.

[4] Note that the 1996 wave of the Bangladesh panel contained four rounds, over which an intra-annual village level attrition rate can be calculated. This will not be feasible for many panel surveys.

```
#delimit ;
xi: probit A $headchar hhsize $demog ownland lvasset96 i.thana
 lpcx96 perfloods9607 villattrate i.categ, robust cluster(village) ;
#delimit cr
```

Table 1: Attrition Probit for Consumption Expenditures

```
Probit regression                            Number of obs   =      954
Log pseudolikelihood = -162.38681             Pseudo R2      =   0.1331
 (Std. Err. adjusted for 47 clusters in village)
```

| Variable (1996 values) | Coefficient | Robust<br>Std Err | z | P-<br>value |
|---|---|---|---|---|
| Age of household head | 0.003 | 0.008 | 0.330 | 0.743 |
| Age of household head squared | 0.001 | 0.000 | 2.040 | 0.042 |
| Education of household head (years) | 0.017 | 0.020 | 0.830 | 0.408 |
| Household size | -0.074 | 0.066 | -1.120 | 0.262 |
| % of household members aged | | | | |
| 0-4 years | 0.013 | 0.006 | 2.090 | 0.037 |
| 5-14 years | 0.010 | 0.005 | 2.030 | 0.042 |
| 15-19 years | 0.003 | 0.006 | 0.490 | 0.621 |
| 35-54 years | 0.015 | 0.007 | 2.230 | 0.026 |
| 55 and older | 0.012 | 0.007 | 1.860 | 0.062 |
| Total land owned (decimals) | 0.000 | 0.001 | -0.030 | 0.977 |
| (Log) Value of Assets | -0.108 | 0.100 | -1.080 | 0.281 |
| Pakundia thana | 0.842 | 0.218 | 3.860 | 0.000 |
| Gaffargao thana | 0.532 | 0.270 | 1.970 | 0.049 |
| Jessore thana | 0.321 | 0.241 | 1.330 | 0.182 |
| (Log) Per Capita Expenditure | -0.128 | 0.219 | -0.580 | 0.560 |
| % of households in village experiencing floods | 0.002 | 0.003 | 0.690 | 0.487 |
| Village Attrition Rate in 1996 | 0.013 | 0.029 | 0.430 | 0.665 |
| Adoption Status in 1996 | | | | |
| B (adoptor, comparison village) | -0.077 | 0.175 | -0.440 | 0.661 |
| C (likely adoptor, comparison) | 0.114 | 0.225 | 0.500 | 0.614 |
| D (non-adopter, comparison village) | 0.373 | 0.208 | 1.790 | 0.073 |
| Constant | -1.977 | 1.340 | -1.480 | 0.140 |

The pseudo R-squared from the attrition probit in Table 1 suggests that baseline variables and village attrition explain about 13% of panel attrition between 1996 and 2006/07. While this is relatively high explanatory power for attrition probit, note that it still leaves some 87% of attrition as unexplained. The z-statistics and P-values in the middle two columns of the table show only six of the 22 variables in the attrition probit are statistically different from zero at the 5% level, although two more are statistically different from zero at the 10% level . Variables that are significant predictors of attrition including the age of the household head squared, selected household demographic variables and one thana (sub-district) and one adoption status dummy.

Using the Stata test command we perform a Wald test of whether these groups variables are jointly equal to zero using the command:

```
#delimit ;
  test  $headchar $demog lpcx96 perfloods9607 villattrate
_Icateg_96_2 _Icateg_96_3 _Icateg_96_4 _Ithana_2 _Ithana_9 _Ithana_
;
#delimit cr
```

The resulting Chi-squared statistic of 85.00 with 17 degree of freedom indicates these variables are jointly statistically different from zero at the highest level of significance (the P-value is 0.000), so we can conclude these are significant predictors of attrition. Notice that the characteristics of the household head and demographic composition variables are among the  variables which are jointly able to predict attrition.  None of the seven groups of variables are, however, individually different from zero at conventional levels of significance.

The BGLW test for attrition is also implemented by creating variables with the interactions between the attrition variable (A) and all other variables using Stata's *xi* command.

```
#delimit ;
xi i.A*agehh i.A*agesqr i.A*educ_h i.A*p0_4 i.A*p5_14 i.A*p15_19 i.A*p35_54
i.A*p55p i.A*ownland i.A*lvasset96 i.A*i.thana
i.A*perfloods9607 i.A*i.categ, prefix(I) ;
#delimit cr
```

A (clustered) regression is then estimated, with the log of per capita expenditures in 1996 as the dependent variable, and the household and auxiliary variables plus their interactions with the Attrition variable (denoted by IAX*) as the explanatory variables:

```
#delimit ;
xi: reg lpcx96 hhsize agehh agesqr educ_h
p0_4 p5_14 p15_19 p35_54 p55p ownland lvasset96 i.thana
perfloods9607  villattrate  i.categ A IAX*, robust cluster(village) ;
#delimit cr
```

Stata's testparm command is then used to test for whether the attrition dummy and all the interactions are jointly equal to zero:

```
#delimit ;
testparm A  IAXagehh_1 IAXagesq_1  IAXeduc__1 IAXp0_4_1 IAXp5_14_1
IAXp15_1_1 IAXp35_5_1 IAXp55p_1
```

```
IAXownla_1 IAXlvass_1 IAXtha_1_2 IAXtha_1_9 IAXtha_1_47 IAXperfl_1
IAXcat_1_2 IAXcat_1_3 IAXcat_1_4 ;
#delimit cr
```

The F-statistic of 24.67 is able to reject the null hypothesis that attrition is random at the highest levels of significance.

Given that both the standard tests indicate that attrition for the expenditure model is non-random, we proceed to calculate inverse probability weights for this model.  To do this we first calculate the predicted probabilities from the unrestricted attrition probit in Table 1, and then re-estimate it excluding seven groups of auxiliary variables, which include the characteristics of the household head, demographic composition and per capita income of the household in the initial period, as well as floods, the thana and treatments dummies, and the treatment dummies.  After calculating the predicted probabilities from the restricted attrition probit, the inverse probability weights are calculated straightforwardly by taking the ratio of the restricted to unrestricted probabilities. These steps are accomplished using the following Stata commands:

```
#delimit ;
xi: probit A $headchar hhsize $demog ownland lvasset96 i.thana
 lpcx96 perfloods9607 villattrate i.categ, robust cluster(village) ;
#delimit cr

gen sample=e(sample)

predict pxav

xi: probit A hhsize ownland lvasset96 if sample==1, robust cluster(village)

predict pxres

gen attwght=pxres/pxav
```

(In the example do-file, *attrition_weights.do*, there are also some additional capture drop commands to ensure that existing predicted values which may be in the memory are deleted).

The inverse probability (or attrition) weights produced vary from .09 to 31.53, with a mean value of 1.48. In general, the inverse probability weights give more weight to households that have remained in the panel than an unweighted regression would, although in some cases households whose characteristics make it very unlikely that they attrit are weighted downwards.

Tables 3 and 4 below show the relatively minor impact that applying these inverse probability weights to standard poverty transition matrices has. Without weighting, households moving out of poverty account for around 50.7% of panel households, while with weighting these households account for 50.5%. Similarly, the number of chronically poor households (households that are poor in both periods) falls from 11.2% without weighting to 9.7% of households with attrition weights.

Table 2: Poverty Transition Matrix Without Attrition Weights

| Poor 1996 | Poor 2007 | | Total |
|---|---|---|---|
| | Poor | Non-Poor | |
| Poor | 11.23 | 50.66 | 61.89 |
| Non-Poor | 1.76 | 36.34 | 38.11 |
| Total | 13.00 | 87.00 | 100.00 |

Table 3: Poverty Transition Matrix With Attrition Weights

| Poor 1996 | Poor 2007 | | Total |
|---|---|---|---|
| | Poor | Non-Poor | |
| Poor | 9.71 | 49.72 | 59.43 |
| Non-Poor | 2.00 | 38.56 | 40.57 |
| Total | 11.72 | 88.28 | 100.00 |

Note that these transition matrices are calculated using the Bangladesh Bureau of Statistics' upper poverty line for 2005 adjusted to 1996 and 2007 terms by the food and non-food consumer price indices.

Expenditure regressions (not shown, but see the code included in attrition_weights.do) show that whether inverse probability weights are applied makes a fairly small difference to semi-log expenditure regressions.

When the value of household assets is the outcome variable of interest, we have a slightly larger sample of 963 households of whom 57 attrit between 1996 and 2007. Table 4 shows the an attrition probit for asset attritors, in which we find limited evidence of non-random attrition, with just three of our independent variables being significantly different from zero at the 5% level. The only one of auxiliary attrition variables which is significantly different from zero at this level is the (natural) logarithm of the value of assets in 1996. A Wald test of the joint significance of the variables related to the household head, the demographic composition of the household, the auxiliary variables and the thana dummies has a Chi-squared value of, 96.92 and a P-value of 0.00, so the null hypothesis of random attrition can be easily rejected. The BGLW pooling test has a F value of 7.35 and a P-value of 0.000, confirming that asset attrition is not random.

Table 4: Attrition Probit for Value of Household Assets

Probit regression Number of obs = 963
Log pseudolikelihood = -188.15118 Pseudo R2 = 0.1306
(Std. Err. adjusted for 47 clusters in village)

| | | Robust | | |
|---|---|---|---|---|
| **Variable (1996 values)** | **Coefficient** | **Std Err** | **z** | **P-value** |
| Age of household head | -0.007 | 0.006 | -1.060 | 0.290 |
| Age of household head squared | 0.001 | 0.000 | 1.950 | 0.051 |
| Education of household head (years) | 0.017 | 0.018 | 0.930 | 0.351 |
| Household size | -0.009 | 0.063 | -0.140 | 0.886 |
| % of household members aged | | | | |
| 0-4 years | 0.009 | 0.007 | 1.240 | 0.215 |
| 5-14 years | 0.005 | 0.005 | 0.970 | 0.330 |
| 15-19 years | -0.001 | 0.007 | -0.140 | 0.889 |
| 35-54 years | 0.011 | 0.007 | 1.670 | 0.094 |
| 55 and older | 0.014 | 0.006 | 2.590 | 0.010 |
| Total land owned (decimals) | 0.000 | 0.001 | 0.770 | 0.442 |
| Attrition Variables | | | | |
| (Log) Value of Assets in 1996 | -0.272 | 0.106 | -2.570 | 0.010 |
| % of households in village experiencing floods | 0.000 | 0.003 | 0.000 | 0.999 |
| Village Attrition Rate in 1996 | 0.036 | 0.030 | 1.180 | 0.237 |
| B (adoptor, comparison village) | -0.213 | 0.187 | -1.140 | 0.256 |
| C (likely adoptor, comparison) | 0.103 | 0.152 | 0.680 | 0.499 |
| D (non-adopter, comparison village) | 0.302 | 0.198 | 1.530 | 0.127 |
| Pakundia thana | 0.646 | 0.237 | 2.730 | 0.006 |
| Gaffargao thana | 0.237 | 0.275 | 0.860 | 0.389 |
| Jessore thana | -0.089 | 0.237 | -0.380 | 0.707 |
| Constant | -1.741 | 0.392 | -4.440 | 0.000 |

Accordingly the groups of variables which predict attrition were dropped and the restricted probit was estimated. The ratios of the predicted values of the restricted to the unrestricted model were calculated, producing inverse probability weights which range from 0.15 to 12.81.

Table 5 report linear panel regressions for the (natural) logarithm of household assets with and without inverse probabilities weights. The regressors are the same basic variables as in attrition probits but exclude the auxiliary variables. These two regressions are produced by the following Stata commands:

```
* without inverse probability weights
#delimit ;
xi: reg lvasset07 $headchar1 hhsize $demog2 $land i.thana, robust
cluster(thana) ;

*with inverse probability weights
xi: reg lvasset07 $headchar1 hhsize $demog2 $land i.thana  [pw=attwght2],
robust cluster(thana);
#delimit cr
```

Table 5: Linear Regressions for Log of Household Assets, 2006-07

| | Without Attrition Weights | With Attrition Weights |
|---|---|---|
| Number of obs = | 906 | 906 |
| R-squared = | 0.273 | 0.254 |
| Root MSE = | 1.1044 | 1.0387 |

(Std. Err. adjusted for 47 clusters in village)

| **Variable**<br>**(1996 values)** | **Without Attrition Weights** | | | **With Attrition Weights** | | |
|---|---|---|---|---|---|---|
| | | **Robust** | | | **Robust** | |
| | **Coefficient** | **Std Err** | **P-value** | **Coefficient** | **Std Err** | **P-value** |
| Age of household head | 0.011 | 0.005 | 0.387 | 0.005 | 0.006 | 0.033 |
| Age of household head squared | 0.000 | 0.000 | 0.168 | -0.001 | 0.000 | 0.622 |
| Education of household head (years) | 0.111 | 0.011 | 0.000 | 0.098 | 0.013 | 0.000 |
| Household size | 0.005 | 0.021 | 1.000 | 0.000 | 0.021 | 0.817 |
| % of household members aged | | | | | | |
| 0-4 years | -0.014 | 0.004 | 0.008 | -0.014 | 0.005 | 0.001 |
| 5-14 years | -0.003 | 0.003 | 0.229 | -0.004 | 0.003 | 0.346 |
| 15-19 years | -0.006 | 0.004 | 0.179 | -0.007 | 0.005 | 0.144 |
| 35-54 years | -0.011 | 0.004 | 0.021 | -0.009 | 0.004 | 0.006 |
| 55 and older | -0.015 | 0.005 | 0.123 | -0.007 | 0.005 | 0.007 |
| Total land owned (decimals) | 0.001 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 |
| Pakundia thana | -0.681 | 0.156 | 0.000 | -0.570 | 0.120 | 0.000 |
| Gaffargao thana | -0.726 | 0.177 | 0.000 | -0.801 | 0.180 | 0.000 |
| Jessore thana | -0.541 | 0.145 | 0.001 | -0.503 | 0.146 | 0.001 |
| Constant | 9.611 | 0.213 | 0.000 | 9.967 | 0.265 | 0.000 |

Inspection of the left and right hand sides of Table 5 reveals that the coefficients and significance of individual coefficients are fairly similar. However, using the usual 5% level of significance, there are two variables (the age of the household head, and the percentage of household members over 55 years of age) which is not significantly different from zero in the regression without attrition weights but significant when weights are applied. A Hausman test of the equality of the coefficients of the weighted and unweighted models is firmly rejected (Chi-squared(13) value of 197.33, P-value of 0.000).

## Conclusions and caveats

Inverse probability weights are one of two methods in common use for correcting for attrition bias (the other being the estimation of a Heckman type selectivity model).
The great advantage of inverse probability weights is their simplicity but a number of caveats must be borne in mind about their use.

First, the tests and adjustment for attrition presented above assume that attrition is based on observables only. When attrition also depends on unobservable factors, as will often be the case, other methods (such as a selectivity model) need to be used.

Second, the attrition tests and adjustments described above are model specific, and must therefore be repeated for each outcome variable of interest.

Third, only one type of attrition considered here (permanent unit non-response) has been considered in this note.  With multiple wave panels some members (in particular when panel units are individuals) may be missing from one wave of a panel only to reappear at a later date.  While inverse probability weights can also be used to adjust for temporary attrition, and also item non-response (when particular questions are not answered), some modifications to their derivation are necessary.

Fourth, while it is possible to correct for attrition bias, it is always wise to try and minimise attrition at the data collection stage. Some useful strategies for reducing attrition in panel data are described in Hill (2001).

Finally, it should be remembered that many significant factors of the poverty experiences of individuals and households are “suppressed” by the construction of panels, although they are informative in their own right. Qualitative and participatory studies, for example, suggest that extreme poverty often leads to the migration of household members, the dissolution of households, and in the most extreme and heart-rending cases, the death of unsupported individuals.

## References

Alderman, H., Behrman, J., Kohler, H.P., Maluccio, J. and Cotts Watkins, S. (2001) 'Attrition in longitudinal household survey data'. *Demographic Research* 5(4): 79-124.

Becketti, S., Gould, W. Lillard, L., and Welch, F. (1988). 'The Panel Study of Income Dynamics after Fourteen Years: An Evaluation'. *Journal of Labor Economics* 6: 472-92.

Outes-Leon, I. and Dercon, S. (2008). 'Survey Attrition and Bias in Young Lives'. Young Lives Technical Note 5. Oxford: University of Oxford.

Falaris, E. (2003). 'The effect of survey attrition in longitudinal surveys: evidence from Peru, Cote D'Ivoire and Vietnam'. *Journal of Development Economics,* 70: 133-158.

Fitzgerald, J., Gottschalk and Moffit, R. (1998). 'An analysis of sample attrition in panel data'. *Journal of Human Resources, 33(2): 251-299.*

Heckman, J, (1979) 'Sample selection basis as a specification error'. *Econometrica*, 47: 153-161.

Hill, Z. (2001). 'Reducing attrition in panel studies in developing countries'. *Young Lives Working Paper 5.*

Mallucio, J. (2004). 'Using quality of interview information to assess non-random attrition bias in developing country panel data'. *Review of Development Economics,* 8(1): 91-109.

Winkels, J. and Withers S. (2000). 'Panel attrition' in Rose, D. (ed) *Researching Social and Economic Change: The Uses of Household Panel Studies*. London: Routledge.

Wooldridge J. (2002). *Econometric Analysis of Cross-sectional and Panel Data*, Cambridge MA: MIT Press.