Attrition in Longitudinal Household Survey Data

Harold Alderman

Jere R. Behrman

Hans-Peter Kohler

John A. Maluccio

Susan Cotts Watkins

Table of Contents

Attrition in Longitudinal Household Survey Data: Some Tests for Three Developing-Country Samples

Harold Alderman [1], Jere R. Behrman [2], Hans-Peter Kohler [3], John A. Maluccio [4], and Susan Cotts Watkins [5]

Abstract

Longitudinal household data can have considerable advantages over much more widely used cross-sectional data for capturing dynamic demographic relationships. However, a disturbing feature of such data is that there is often substantial attrition and this may make the interpretation of estimates problematic. Such attrition may be particularly severe where there is considerable migration between rural and urban areas. Many analysts share the intuition that attrition is likely to be selective on characteristics such as schooling and thus that high attrition is likely to bias estimates. This paper considers the extent and implications of attrition for three longitudinal household surveys from Bolivia, Kenya, and South Africa that report very high per-year attrition rates between survey rounds. Our estimates indicate that: (a) the means for a number of critical outcome and family background variables differ significantly between those who are lost to follow-up and those who are re-interviewed; (b) a number of family background variables are significant predictors of attrition; but (c) nevertheless, the coefficient estimates for standard family background variables in regressions and probit equations for a majority of the outcome variables considered in all three data sets are not affected significantly by attrition. Therefore, attrition apparently is <u>not</u> a general problem for obtaining consistent estimates

1   Development Research Group, World Bank, 1818 H Street NW, Washington D.C. 20433, USA. Email: halderman@worldbank.org.

2   Population Studies Center, McNeil 160, 3718 Locust Walk, University of Pennsylvania, Philadelphia, PA 19104-6297, USA. Email: jbehrman@econ.sas.upenn.edu.

3   Max-Planck Institute for Demographic Research, Doberaner Str. 114, 18057 Rostock, Germany. Email: kohler@demogr.mpg.de.

4   International Food Policy Research Institute, 2033 K Street NW, Washington D.C. 20006, USA. Email: j.maluccio@cgiar.org.

5   University of Pennsylvania, McNeil 113, 3718 Locust Walk, Philadelphia, PA 19104-6299, USA. Email: swatkins@pop.upenn.edu.

of the coefficients of interest for most of these outcomes. These results, which are very similar to those for developed countries, suggest that multivariate estimates of behavioral relations may not be biased due to attrition and thus support the collection of longitudinal data.

1. Introduction

Longitudinal (or panel) household data can have considerable advantages over more widely available cross-sectional data for social science analysis. Longitudinal data permit (1) tracing the dynamics of behaviors, (2) identifying the influence of past behaviors on current behaviors, and (3) controlling for unobserved fixed characteristics in the investigation of the effect of time-varying exogenous variables on endogenous behaviors. These advantages are substantial for demographers studying processes that occur over time including the impact of programs on subsequent behavior that often use time-varying exogenous variables. As a result, the advantages are also increasingly appreciated: for example, a review of articles published in the journal *Demography* indicates that only 26 articles using longitudinal data appeared between 1980-1989, while there were 65 between 1990-2000.

Unfortunately, the collection of longitudinal data is likely to be difficult and expensive, and some researchers, such as Ashenfelter, Deaton, and Solon (1986), have questioned whether the gains are worth the costs. One problem in particular that has concerned analysts is that sample attrition may lead to selective samples and make the interpretation of estimates problematic. Many analysts share the intuition that attrition is likely to be selective on characteristics such as schooling and thus that high attrition is likely to bias estimates made from longitudinal data. While there has been some work on the effect of attrition on estimates using developed-country samples, little has been done using data from developing countries, where considerable migration between rural and urban areas typically exacerbates the problem of attrition. Table 1 summarizes the attrition rates in a number of longitudinal data sets from developing countries. While these vary widely (ranging from 6 to 50 percent between two survey rounds and 1.5 to 23.2 percent per year between survey rounds), often there is considerable attrition.

In this paper, we consider some of the implications of attrition for three of the seven longitudinal household surveys from developing countries in Table 1 that report the highest per-year attrition rates between survey rounds: (1) a Bolivian household survey designed to evaluate an early childhood development intervention in poor urban areas, with survey rounds in 1995/1996 and 1998; (2) a Kenyan rural household survey designed to investigate the role of social networks in attitudes and behavior regarding reproductive health, with survey rounds in 1994/1995 and 1996/1997; and (3) a South African (KwaZulu-Natal

Table 1:   *Attrition rates for longitudinal household survey data in developing countries listed in order of attrition rates per year*

Country, time period/interval between rounds (in rough order of attrition rates per year)	Attrition rate between rounds (percentage)	Attrition rate per year (percentage)	Source
Bolivia (urban), 1995/6 to 1998 (two-year interval)	35	19.4	Present study (also see Alderman and Behrman 1999)
Kenya (rural, South Nyanza Province), 1994/5 to 1996/7 (two-year interval)			Present study (also see Behrman, Kohler, and Watkins 2001)
couples	41	23.2	
men	33	18.1	
women	28	15.1	
Nigeria (five-year interval)	50	13.0	Renne (1997)
South Africa (KwaZulu-Natal) 1993 to 1998. (five year interval)			Present study (also see Maluccio 2001)
households	16	3.4	
preschool children	22	4.8	
India (rural) 1970/71 to 1981/2 (11-year interval)	33	3.6	Foster and Rosenzweig 1995
Malaysia (12-year interval)	25	2.4	Smith and Thomas 1997
Indonesia 1993 to 1997 (four-year interval)	6	1.5	Thomas, Frankenberg, and Smith 1999

Note: The annual attrition rate is calculated as $1-(1-q)^{1/T}$, where q is the overall attrition rate and T is the number of years covered by the panel.

Province) rural and urban household survey designed for more general purposes, with survey rounds in 1993 and 1998. The different aims of the projects and the variety of outcome measures facilitate generalization, at least for survey areas such as these that are relatively poor and experiencing considerable mobility**.**

Drawing on recent studies on attrition in longitudinal surveys for developed countries, the next section summarizes theoretical aspects of the effects of attrition on estimates. Section 3 describes the three datasets used in this study and section 4 presents some tests for the implications of attrition between the first and the second rounds of the three surveys. Section 5 summarizes our conclusions.

2. Some Theoretical Aspects of the Effects of Attrition on Estimates

Most of the previous work on attrition in large longitudinal samples is for developed economies, for example, the studies published in a special issue of *The Journal of Human Resources* (Spring 1998) on Attrition in Longitudinal Surveys (for related statistical literature on missing values and survey non-response see for instance Little and Rubin 1987 or Ahlo 1990). The striking result of the studies presented in the *Journal of Human Resources* (JHR) is that the biases in estimated socioeconomic relations due to attrition are small despite attrition rates as high as 50 percent and significant differences between those re-interviewed and those lost to follow-up for many important characteristics. For example, Fitzgerald, Gottschalk and Moffitt (1998) summarize:

By 1989 the Michigan Panel Study on Income Dynamics (PSID) had experienced approximately 50 percent sample loss from cumulative attrition from its initial 1968 membership… (p. 251)

> We find that while the PSID has been highly selective on many important variables of interest, including those ordinarily regarded as outcome variables, attrition bias nevertheless remains quite small in magnitude. … (most attrition is random)... (p. 252)

> Although a sample loss as high as [experienced] must necessarily reduce precision of estimation, there is no necessary relationship between the size of the sample loss from attrition and the existence or magnitude of attrition bias. Even a large amount of attrition causes no bias if it is 'random' … (p. 256)

The other studies in this special issue of the *JHR* further confirm these findings for the PSID or reach similar conclusions for other important panel data such as the Survey of

Income and Program Participation (SIPP), the National Longitudinal Surveys of Labor Market Experience (NLS), and the Labor Supply Panel Survey in the Netherlands (Falaris and Peters 1998; Lillard and Panis1998; Van den Berg and Lindeboom 1998; Zabel 1998; Ziliak and Kniesner 1998).

This absence of relevant distortions in parameter estimates due to attrition can be understood once the relation between the mechanisms leading to attrition and the empirical model of interest is made explicit.

2.1 Attrition bias due to selection on observables and unobservables

Fitzgerald, Gottschalk, and Moffitt (1998) provide an econometric framework for the analysis of attrition in which the common distinction between selection on variables observed in the data and variables that are unobserved is used to develop tests for attrition bias and correction factors to eliminate it. (Note 1) This framework assumes a panel study that attempts to interview the same sample of respondents (or households, etc.) for say, T annual survey rounds at times $t = 1, \ldots T$. The initial sample at time $t=1$ is assumed to be a random or stratified random sample of the population. Attrition of a respondent at time t, denoted A_t, is then defined as the fact that the respondent participates in all survey waves $1, \ldots, t\text{-}1$, but does not participate in any survey wave from time t onwards (Note 2). Common causes for attrition are death or migration of the respondent, or refusal to participate due to saturation or frustration with a particular survey. The respondent thus reports information for the dependent and explanatory variables for the survey waves $1, \ldots, t\text{-}1$. Neither the dependent variable nor time-varying explanatory variables are observed from survey wave t onwards. (Note 3) Analyses of and adjustments for attrition at time t can therefore be based on fixed characteristics of the respondent, lagged time-varying variables pertaining to periods prior to time t, and information that do not require the completion of an interview, such as interviewer characteristics and location of residence.

The central concern in the analyses of attrition – and of missing data in general – is selection bias, that is, a distortion of the estimation results due to non-random patterns of attrition. The common distinction is between attrition that is completely random, attrition that is selective on variables unobserved in the data, and attrition that is selective on variables observed in the data. The latter can be further distinguished between attrition that leads to ignorable selection on observables (the statistical literature on missing data also uses the terms "missing-at-random") or non-ignorable selection on observables.

While attrition does not necessarily introduce bias in the estimates of interest, when it does, selective attrition on observables is more amenable to statistical solutions than selective attrition on unobservables. In particular, the above taxonomy of attrition leads to a sequence of tests that we will follow in this study. *First*, given that there is sample

attrition, one determines whether or not there is selection on observables. *Second*, if there is selection on observables, one determines whether this attrition is ignorable – and thus does not bias the estimates of interest – or whether it is non-ignorable. In the latter case, the analyses need to adjust for attrition since otherwise selection leads to biased inferences about relevant parameters. The available methods to correct for attrition on observables are often relatively easy to implement and rely on relatively weak assumptions, in contrast to the methods that are required in order to adjust for selection on unobservables. While selective attrition on unobservables potentially remains a problem even after the analyses account for selection on observables, using as much information as possible about selection on observables in the panel helps to reduce the amount of residual, unexplained variation in the data due to attrition. Controlling for selection on observables thus will likely reduce the biases due to the selection on unobservables. (Note 4)

More formally, consider the survey wave at time t and assume that what is of interest is a conditional population density $f(y_t|x_t)$ where y_t is a scalar dependent variable and x_t is an observed scalar independent variable (for illustration; in practice the extension treating x_t as a vector, which potentially includes lagged dependent variables, fixed characteristics of the respondent, and lagged time-varying characteristics of the respondent, is straightforward; see for instance Fitzgerald et al. 1998). In particular, we assume the linear parametric model

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t,$$
$$y_t \text{ observed if } A_t = 0 \tag{1}$$

where ε_t is a mean-zero random variable, and A_t is an attrition indicator equal to 1 if an observation is missing its value of y_t because of attrition, and equal to zero if an observation is not missing its value of y_t. For identification, we assume in this theoretical model that the variable x_t is *observed* for both attritors and non-attritors, as would be the case if it were a time-invariant or lagged variable, for example. The presence of attrition implies that Eq. (1) can only be estimated for respondents that are interviewed at time t, that is for observations for which $A_t=0$ and y_t is observed.

The analysis of these observed data can therefore determine the density $f(y_t|x_t, A_t=0)$ that is conditional on x_t and $A_t=0$. Additional information or restrictions are necessary in order to infer the density of primary interest, $f(y_t|x_t)$, from the observed data. That is, we seek $f(y_t)$ conditional on x_t but not on $A_t=0$.

This additional information can come from the probability of attrition, $\Pr(A_t=0|y_t, x_t, z_t)$, where z_t is an auxiliary variable (or vector) that is assumed to be observable for all units but is not included in x_t. In particular, in the straightforward generalization to vectors, z_t can include lagged values of the dependent variable (which are observed up to time t-1 for respondents who are lost to follow-up at time t), as well as fixed characteristics of the

respondent, lagged time-varying characteristics, and variables that do not require the completion of an interview, such as interviewer characteristics and location of residence. (The set of respondent characteristics that can potentially be included in z_t is restricted to those characteristics that are not already included among the variables in x_t.)

Linearizing the probability of attrition implies a process of the form

$$A_t^* = \delta_0 + \delta_1 x_t + \delta_2 z_t + v_t \tag{2}$$

$$
\begin{aligned}
A_t &= 1 \text{ if } A_t^* \geq 0 \\
&= 0 \text{ if } A_t^* < 0,
\end{aligned} \tag{3}
$$

where A_t^* is a latent index and attrition occurs if this index is equal or larger to zero and v_t is a mean-zero random influence on the attrition probability.

Attrition can then be classified as follows (this classification differs slightly from that proposed by Fitzgerald et al. 1998 and has a more direct relation to the statistical literature on missing data; see also Kohler 2001):

Attrition exhibits *selection on unobservables* if $\Pr(A_t{=}0|y_t, x_t, z_t) \neq \Pr(A_t{=}0|x_t, z_t)$, so that the attrition function cannot be reduced from $\Pr(A_t{=}0|y_t, x_t, z_t)$. In the specific parametric model in Eqs. (1 – 3), therefore, selection on unobservables occurs if v_t is not independent of $\varepsilon_t|x_t$, where $\varepsilon_t|x_t$ is a shorthand notation for the error term ε_t conditional on x_t.

Attrition exhibits *selection on observables* if

$$\Pr(A_t{=}0|y_t, x_t, z_t) = \Pr(A_t{=}0|x_t, z_t), \tag{4}$$

that is, if, conditional on x_t and z_t, the attrition probability is independent of the dependent variable y_t and therefore of the unobserved factors entering the error term ε_t in relation (1).

On one hand, this selection on observables is *ignorable* if (a) y_t and z_t are independent conditional on x_t and $A_t{=}0$, or (b) the attrition function in Eq. (4) can be further reduced to $\Pr(A_t{=}0|x_t, z_t) = \Pr(A_t{=}0|x_t)$, i.e., the probability of attrition is independent of the variable z_t. Ignorable selection on observables implies that the linear regression of relation (1) on the basis of the observed data on non-attritors leads to unbiased estimates of the coefficients β_0 and β_1. In this case, no specific methods are required to control or adjust for attrition.

On the other hand, selection on observables is *non-ignorable* when neither condition (a) nor (b) holds. In this case, standard linear regression analysis of relation (1) does *not* yield unbiased estimates of the coefficients β_0 and β_1, and alternative estimation techniques are required that are further discussed below. Stated in terms of the parametric model in Eqs. (1 – 3), ignorable selection on observables occurs if v_t is independent of $\varepsilon_t|x_t$ and (a') z_t is independent of $\varepsilon_t|x_t$, or (b') the attrition does not depend on z_t (i.e., δ_2 in Eq. 2 is zero).

Selection on observables in this parametric model is non-ignorable when neither condition (a') nor (b') holds.

Attrition is *completely at random* if the attrition function $\Pr(A_t=0|y_t, x_t, z_t)$ can be reduced to $\Pr(A_t=0)$ and attrition neither depends on the dependent variable y_t nor the observed variables x_t and z_t. In our specific model, attrition is completely at random if v_t is independent of $\varepsilon_t|x_t$ and δ_1 and δ_2 in Eq. (2) are zero.

Ordering these attrition patterns in terms of their assumptions from more restrictive to less restrictive yields: completely random attrition < selective attrition on observables < selective attrition on unobservables. Completely random attrition is unlikely in most panel studies, and if it exists, it does not result in biases of parameter estimates. Attrition that is selective on observables and unobservables, on the other hand, is probably a common phenomenon in most panel studies, and we will briefly discuss the statistical approaches to overcome the biases that are potentially caused by such attrition.

Selection on unobservables is often presented as dependent on the estimation of the attrition index equation (2) (see for instance Maddala 1983 or Powell 1994 for discussions of this approach). Identification, however, usually relies on nonlinearities in the index equation or an exclusion restriction, i.e., the existence of a variable z_t – often loosely termed "instrument" – that predicts attrition but is independent of $\varepsilon_t|x_t$ and not included in x_t. It is difficult to rationalize most such exclusion restrictions because, for example, personal characteristics that affect attrition might also directly affect the outcome variable, i.e., they should be in x_t or are correlated with $\varepsilon_t|x_t$. There may be some such identifying variables in the form of variables that are external to individuals and not under their control, such as characteristics of the interviewer in the various rounds (Zabel 1998, Maluccio 2001). However, in the PSID and potentially also in other panel studies the interviewers are assigned on the basis of respondent characteristics, in which case this strategy is also not feasible. In general, therefore, selection on unobservables presents an obstacle to accurate parameter estimation. Most promising, in our opinion, is therefore to test and – if necessary adjust – for non-ignorable selection on observables by using as much information as possible about selection in the panel. This reduces the amount of residual, unexplained variation due to attrition left over in the data and it lessens the scope for selection on unobservables for which few feasible statistical solutions exist.

If there is non-ignorable selection on observables, the critical variable is z_t, a variable that affects attrition propensities and that is also related to the density of y_t conditional on x_t due to the fact that z_t is not independent of $\varepsilon_t|x_t$. In this sense, z_t is "endogenous to y_t." Indeed, a lagged value of y_t can play the role of z_t if it is not in the structural relation being estimated but is related to attrition.

Fitzgerald et al. (1998) show formally that, under the selection on observables restriction in Eq. (4), the complete population density $f(y_t|x_t)$ can be computed from the conditional joint density of y_t and z_t, which we denote by g:

$$f(y_t|x_t) = \int g(y_t, z_t \mid x_t, A_t=0) \; w(z_t, x_t) \; dz_t, \tag{5}$$

where

$$w(z_t, x_t) = \Pr(A_t=0|x_t) \,/\, \Pr(A_t=0|z_t, x_t) \tag{6}$$

are normalized weights (the proof of Eq. 5 is also given in the appendix of this paper). (Note 5) The numerator of Eq. (6) is the probability of remaining in the sample (i.e., non-attrition) conditional on x_t, and the denominator is the probability of remaining in the sample conditional on z_t and x_t. The weights $w(z_t, x_t)$ in Eq. (6) can be estimated from the data when both x_t and z_t are observed. This is the case when – as we have assumed above – x_t and z_t contain either time-invariant or lagged time-varying characteristics of the respondent or variables that do not require a completed interview. (Note 6)

The intuition for Eqs. (5 – 6) is in the spirit of weighting (panel) observations with the inverse of the probability that an observation is included (as in stratified samples, for instance); in the above case pertaining to attrition, this probability is replaced by the function of attrition probabilities in Eq. (6). Because both the weights and the conditional density g are identifiable and estimable from the data, the complete-population density $f(y_t|x_t)$ is estimable as well as its moments such as the expected value $Ey_t = \beta_0 + \beta_1 x_t$ implied by Eq. (1). This result is particularly important since it implies that in the linear model in Eq. (1) the parameters β_0 and β_1 can be estimated without bias, despite the presence of selective attrition on observables, via a weighted least squares regression (WLS) that uses the weights defined in Eq. (6).

Inspection of Eqs. (5) and (6) also reveals the cases when selection on observables can be ignored. In particular, if z_t is not a determinant of attrition, the weights in Eq. (6) equal one and no attrition bias is present. If y_t and z_t are independent conditional on x_t and $A_t=0$, the density g in Eq. (5) factors and it can again be shown that the unconditional density $f(y_t|x_t)$ equals the conditional density and there is no attrition bias.

2.2 Testing for attrition bias (Note 7)

Testing for attrition bias due to selection on unobservables is possible in econometric models that include the estimation of the attrition index. The identification of such models with panel data, however, is problematic due to the frequent lack of instruments that allow identification. As an alternative, Fitzgerald et al. (1998) suggest that indirect tests for selection on unobservables can be made by comparisons with data sets without (or with much less) attrition (e.g., the Current Population Survey for comparison with the PSID in

the United States). Unfortunately, only very limited possibilities for such comparisons exist for most panels, and such comparisons are especially difficult in developing countries. Due to this limited ability to detect selective attrition on unobservables with the datasets examined in this paper, we do not discuss this approach further nor do we perform the corresponding tests.

Testing for selection bias due to selective attrition on observables, on the other hand, is possible in most panel studies and we will focus on these approaches. The two sufficient conditions that render the selection on observables through attrition *ignorable* are either (1) z_t does not affect A_t or (2) z_t is independent of y_t conditional on x_t and $A_t=0$. Specification tests can be based on either of these two conditions. One test is simply to determine whether candidate variables for z_t (for example, lagged values of y) significantly affect A_t. Another test is based on Becketti, Gould, Lillard, and Welch (1988). In the BGLW test, the value of y at the initial wave of the survey (y_1) is regressed on respondent's characteristics at the initial wave (x_1) and on A, which denotes the event that a respondent becomes an attritor at some time during the survey (i.e., A_t equals one for some t in $2,\ldots,T$). The test for attrition is based on the significance of A in that equation. This test is closely related to the test based on regressing A on x_1 and y_1, which is a direct estimation of the attrition probability in Eqs. $(2-3)$ in the special case when the y_1 is used to represent the auxiliary variable z_t. In fact, the direct estimation of the attrition probability and the BGLW test are simply inverses of one another (Fitzgerald et al. 1998). (Note 8)

Clearly, if there is no evidence of attrition bias from these specification tests, this suggests that the attrition on observables is ignorable. (Since the null-hypothesis of our attrition tests is the absence of attrition, the fact that there is not significant evidence of attrition bias from these specification tests is no proof that such bias does not exist. It does, however, show that the possible bias is too small to be detectable given the power of the available tests. This limitation is a general problem of statistical inference and not restricted to the specification tests for attrition).

If the specification tests suggest that attrition on observables is ignorable, then the desired information on $f(y_t|x_t)$ can be directly inferred from the conditional density $f(y_t|x_t, A_t=0)$ (under the assumption that there is no selective attrition on unobservables). If the above tests detect non-ignorable selection on observables due to attrition, the resulting biases in the inference of β_0 and β_1 in Eq. (1) can be avoided by using a weighted least squares methodology with the weights given in Eq. (6).

3. Data and Extent of Attrition

In this section, we describe the three data sets that we use, emphasizing the diverse relations of interest they can address.

3.1 Bolivian Pre-School Program Evaluation Household Survey Data. El Proyecto Integral de Desarrollo Infantil (PIDI)

PIDI is a targeted urban early child development project expected to improve the nutritional status and cognitive development of children who participate and to facilitate the labor force participation of their caregivers. PIDI delivers child services through childcare centers located in the homes of local women who have been trained in childcare. The program provides food accounting for 70 percent of the children's nutritional needs, health and nutrition monitoring, and programs to stimulate the children's social and intellectual development. The PIDI program was designed to facilitate ongoing impact evaluation through the collection of longitudinal data.

Eligibility for PIDI at the time of the collection of the first and second rounds of data was based on an assessment of social risk. As a result of this selection, children who attend a PIDI center are, on average, from poorer family backgrounds than children who live in the same communities but who do not attend a PIDI center (Behrman, Cheng and Todd 2001). The first PIDI evaluation data set (Bolivia 1) was collected between November 1995 and May 1996 and consisted of 2,047 households. (Note 9) The follow-up survey (Bolivia 2) was collected in the first half of 1998 and consisted of interviews in the 65 percent of the original 2,047 households that could be located (plus an additional 3,453 households that were not visited in Bolivia 1). The attrition rate of 35 percent for Bolivia 1 is relatively high, which raised concern about whether reliable inferences could be drawn from analysis of Bolivia 2.

3.2 The Kenyan Ideational Change Survey (KDICP)

KDICP is a longitudinal survey designed to collect information for the analysis of the roles of informal networks in understanding change in knowledge and behavior related to contraceptive use and prevention of AIDS. Four rural sites (sublocations) were chosen in Nyanza Province, near Lake Victoria in the southwestern part of Kenya. The sites were chosen to be similar in most respects but to maximize variation along two dimensions: 1) the extent to which social networks were confined to the sublocation versus being geographically extended and 2) the presence or absence of a community-based distribution program aimed at increasing the use of family planning. Villages were selected randomly within each site and interviews were attempted with all ever-married women of childbearing age (15 – 49) and their husbands. The study consisted of ethnographic interviews, focus groups, and a household survey of approximately 900 women of reproductive age and their husbands, and was conducted between December 1994 and January 1995 (Kenya 1). A second round was conducted in 1996/1997 (Kenya 2). (The surveys are described in detail

at www.pop.upenn.edu/networks). The attrition rates between the two surveys were 33 percent for men, 28 percent for women, and 41 percent for couples (Table 1). (Note 10) These rates are comparable to the 35 percent reported for the Bolivian data.

Table 2 summarizes data on the reported causes of attrition for men and women as obtained from other household members for most individuals who were interviewed in Kenya 1 but not in Kenya 2. (Note 11) Nyanza Province has a relatively high level of AIDS: mortality between the surveys accounted for 18 percent of the reasons given for men's attrition, but only half as much (10 percent) for women. For both men and women the leading explanation was migration, accounting for 59 percent of the reasons given for women and 48 percent of the reasons given for men. Because this is a patrilocal society, a significant share of this migration (over one-third) for women was associated with divorce or separation, but this was not a major factor for men. Not being found at home after at least three visits by interviewers was the next most common explanation for attrition in Kenya 2, accounting for about one-sixth of the reasons given for both men (18 percent) and women (16 percent). Explicitly refusing or claiming to be too busy or sick to participate accounted for slightly smaller percentages – 16 percent for men and 11 percent for women (with most of this gender difference accounted by "other," which is 4 percent for women but 0 percent for men).

3.3 KwaZulu-Natal Income Dynamics Study (KIDS)

The first South African national household survey, the 1993 Project for Statistics on Living Standards and Development (PSLSD), was undertaken in the last half of 1993 under the leadership of the South African Labour and Development Research Unit (SALDRU) at the University of Cape Town. (Note 12) This analysis uses a subset of these data comprising Africans and Indians living in KwaZulu-Natal Province and described further below. Unlike the special purpose household surveys for Bolivia and Kenya, the South African survey was a comprehensive household survey similar to a Living Standards Measurement Survey (Grosh and Glewwe 2000) and collected a broad array of socioeconomic information from individuals and households. Among other things, it included sections on household demographics, household environment, education, food and nonfood expenditures, remittances, employment and income, agricultural activities, health, and anthropometry (weights and heights of children aged six and under). The 1993 sample was selected using a two-stage, self-weighting design. In the first stage, clusters were chosen proportional to population size from census enumerator districts or approximate equivalents when these were unavailable. In the second stage, all households in each chosen cluster were enumerated and then a random sample selected (see PSLSD 1994 for further details).

Table 2: *Reported reasons for men's and women's attrition in Kenyan (KDICP) survey*

	Men		Women	
Reason for attrition:	Number	Percentage	Number	Percentage
Working, moved to, or visiting outside Nyanza Province	45	22.4	21	10.3
Working, moved to, or visiting elsewhere in Nyanza Province	51	25.4	56	27.6
Not home	36	17.9	32	15.8
Refused	26	12.9	20	9.9
Sick or busy	6	3.0	3	1.5
Deceased	37	18.4	20	9.9
Separated, divorced, then moved away	n/a	n/a	42	20.7
Other	0	0.0	11	4.4
Total	201		205	

Note: n/a = not available

Since the 1993 survey, South Africa has undergone dramatic political, social, and economic change, beginning with the change of government after the first national democratic elections in 1994. With the aim of addressing a variety of policy research questions concerning how individuals and households were faring under this transition, African and Indian households surveyed by the PSLSD in South Africa's most populous province, KwaZulu-Natal, were resurveyed from March to June, 1998, for the KIDS (see May et al. 2000). In this paper, the sample of 1993 PSLSD African and Indian households residing in KwaZulu-Natal is referred to as South Africa 1 and those re-interviewed in 1998 for the KIDS, South Africa 2.

An important aspect of the South Africa resurvey – differentiating it further from the Bolivian and Kenyan longitudinal surveys – is that, when possible, the interviewer teams tracked, followed, and re-interviewed households that had moved. (Note 13) Hence, in the South Africa survey migration does not imply automatic attrition from the sample. In addition to reducing the level of attrition and allowing analysis of migration behavior, tracking and following plausibly reduced biases introduced by attrition, a claim we evaluate below.

In 1993, the KwaZulu-Natal sample contained 1,354 households (215 Indian and 1,139 African). Of the target sample, 1,152 households (84 percent) with at least one 1993 member were successfully re-interviewed in 1998 (Maluccio 2001). As in most surveys in developing countries, refusal rates were very low, less than 1 percent. The remaining households that could not be re-interviewed were either verified as having moved but could not be tracked (7 percent) or left no trace (8 percent). Had the sixty households that had moved but were successfully tracked not been followed, 79 percent of the target households would have been re-interviewed. Put another way, the tracking procedures yielded a 25 percent reduction in the number of households that were lost to follow-up.

Re-interview rates were slightly higher in urban than in rural areas. Offsetting that success was a follow-up rate of 78 percent (of 215 households) for Indian households, all of which were urban. The follow-up rate for rural Africans was 83 percent (of 825 households). There were no major differences in the analysis of attrition when we considered the rural and urban samples separately; therefore we present only the results where we pooled them.

The discussion of attrition between South Africa 1 and South Africa 2 to this point has focused on attrition at the household level. For an analysis of individual level outcomes, however, attrition at the individual level is the relevant measure. Because a household was considered to be found if at least one 1993 member was re-interviewed, individual-level attrition for the entire sample is necessarily higher than household attrition (although this need not be the case for subsamples of individuals). Focusing on the sample of children aged 6 – 72 months for whom there is complete information on height, weight, and age in 1993, for example, 78 percent of 897 children were re-interviewed as household members in 1998, indicating one-third more attrition than at the household level. (Note 14).

4. Some Attrition Tests for the Bolivian, Kenyan, and South African Samples

As noted, the attrition rates for the three samples considered here are considerable: 35 percent for the Bolivian sample, from 28 percent for women to 41 percent for couples in the Kenyan sample, and from 16 percent for households to 22 percent for pre-school

children in the South African sample. However, studies for developed countries suggest that while attrition of this magnitude may be selective, it need not significantly affect estimated multivariate relations. To test this, we conducted three sets of tests of attrition as it relates to observed variables in the data, using some of the tests presented by Fitzgerald, Gottschalk, and Moffitt (1998). We begin with a comparison of means, since the intuition that attrition is likely to bias estimates is often made on the basis of such univariate comparisons. We then estimate probits for the probability of attrition in order to ask what variables predict attrition comparing univariate and multivariate estimates. Lastly, we test whether coefficient estimates for a set of relations of interest to the objectives of the studies differ for two subsamples, one that is lost to follow-up and one that is re-interviewed.

4.1 Comparison of Means for Major Outcome and Control Variables

First, we compared means for major outcome and control variables measured in the first rounds of the respective data sets for those subsequently lost to follow-up versus those who were re-interviewed (Tables 3, 4, and 5). Major characteristics are defined with respect to the interests of the project for which these data were collected.

Bolivia: A number of means for those lost to follow-up differ statistically from those who eventually were re-interviewed: rates of severe stunting, moderate wasting, the fraction reporting that they mainly spoke Quechua at home, weight-for-age, gross motor ability test scores, fine motor ability test scores, language-audition test scores, personal-social test scores, mother's age, father's age, home ownership, fraction with both parents present, number of rooms in the home, number of siblings, ownership of durables, mother having job, and household income (Table 3). All of these observable characteristics distinguish the two subsamples at least at the 10 percent significance level, and show that in the first round of the data (Bolivia 1) children who were worse off in terms of these measures were more likely to be lost to follow-up before the second round than those who would eventually be re-interviewed. Among the fourteen predetermined parental and household level variables in Table 3, eleven differ significantly for the two groups at least at the 10 percent significance level. Thus, both in terms of child development outcome variables and family background variables, attrition seems to be systematically more likely for children who are worse off. Such systematic differences, together with the high attrition rates, may cause concern about what can be inferred with confidence from these longitudinal data.

Kenya: For the Kenyan data, both males and females lost to follow-up have higher schooling, more languages, and are more likely to have heard radio messages about contraception and lived in households with males who received salaries (Table 4). They are also younger and have fewer children than those who were re-interviewed. For a few variables the means differ significantly between these two subsamples for men but not for

women (ever-use of contraceptives, residence in the sublocation of Owich) or for women but not for men (want no more children, visited by community-based distribution agent, speaks Luo only, belongs to credit group or to clan welfare society, residence in the sublocation of Wakula South). On the other hand, the means do not differ for the subsamples of either men or women for a number of characteristics (currently using contraceptives, heard about family planning at clinic, discussed family planning with others, number of partners in networks, primary schooling, lived outside of province, polygamous household).

Therefore, it appears that attrition is selective in terms of some 'modern' characteristics (including some of the outcome variables that these data were designed to analyze) with selectivity more strongly related to women's characteristics. But the means for many characteristics, including those for most of the indicators of social interaction, the impact of which is central to the project for which these data were gathered, do not differ significantly between those lost to follow-up and those re-interviewed.

South Africa: Because the South African survey is a comprehensive household survey with a large number of variables, for comparability this study examined a set of variables similar to those considered for Bolivia, i.e., measures of child nutritional status based on anthropometrics, as well as a set of predetermined family background characteristics. The results reported here cannot, therefore, be immediately generalized to other outcome variables available in the South African data.

There are no significant differences in the means of child nutritional status outcome variables between the two groups (Table 5). This is not the case for the predetermined family background variables, however, where there are a number of significant differences at the ten percent level of significance. Those pre-school children who were re-interviewed are significantly more likely to be African rather than Indian, and come from households that have lower income, less educated heads, and fewer durable assets. Of course, since these background variables themselves tend to be highly correlated (in particular race with income and assets), it is not surprising that they show similar patterns in the comparisons of means. Households residing in the former Natal Province areas of the province were also less likely to be re-interviewed; this likely reflects higher migration, in part due to weaker property rights, in those areas. In sum, while there are no apparent differences in the child outcome variables, children from better off or Indian households were more likely to be lost to follow-up.

Table 3: *Bolivia. T-tests for differences in means in Bolivia 1 data for attritors versus nonattritors* [a]

Variables	Re-interviewed		Not re-interviewed		Difference	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	t-test
Early child development outcome variables						
Height-for-age[b]	18.0	(22.5)	17.4	(22.1)	0.65	(0.72)
Weight-for-age[b]	32.2	(26.5)	30.3	(25.8)	1.91*	(1.81)
Weight-for-height[b]	58.1	(26.5)	56.9	(27.2)	1.21	(1.10)
Moderate stunting[c]	0.639	(0.48)	0.631.	(0.48)	0.008	(0.43)
Severe stunting[c]	0.279	(0.45)	0.323	(0.47)	-0.0437**	(-2.37)
Moderate wasting[c]	0.365	(0.48)	0.400)	(0.49	-0.035*	(-1.79)
Severe wasting[c]	0.0796	(0.27)	0.0946	(0.29)	-0.0150	(-1.30)
Gross motor ability	20.8	(7.81)	20.3	(7.67)	0.5136*	(1.65)
Fine motor ability	19.4	(7.28)	19.0	(7.19)	0.480*	(1.65)
Language-audition	19.2	(7.62)	18.6	(7.44)	0.569*	(1.88)
Personal-social	19.9	(8.02)	19.4	(8.06)	0.534*	(1.65)
Predetermined family background variables						
Mother's age	29.8	(6.45)	28.7	(6.44)	1.07**	(4.10)
Father's age	33.0	(7.70)	32.2	(8.03)	0.85**	(2.66)
Mother's schooling	3.0	(1.5)	3.0	(1.5)	-0.06	(-0.9113)
Father's schooling	3.6	(1.4)	3.6	(1.4)	-0.02	(-0.42)
Quechua mainly	.00099	(0.0315)	0.0114	(0.106)	-0.00414**	(-2.85)
Amarya mainly	.00396	(0.0628)	0.00456	(0.07)	-0.000605	(-0.23)
Home ownership	0.428	(0.495)	0.215	(0.411)	0.213**	(12.02)
Number of rooms in house	1.50	(1.05)	1.40	(1.00)	0.100**	(4.17)
Both parents present	0.841	(0.366)	0.775	(0.42)	0.0656**	(4.54)
Number of siblings	2.37	(1.80)	2.05	(1.59)	0.322**	(4.80)
Ownership of durables[d]	6.30	(2.11)	5.92	(1.92)	0.375**	(4.69)
Job of mother[e]	2.26	(0.91)	2.08	(0.91)	0.174**	(4.73)
Job of father	2.70	(0.54)	2.70	(0.55)	-0.006	(-0.28)
Household income	922	(755)	868	(638)	54**	(2.68)

Notes: * indicates significance at the 10 percent level, and ** at the 5 percent level.

[a] Values of two-sample t-test with unequal variances are given in parentheses in last column.

[b] Height-for-age in centimeter/years. Weight-for-age in kilogram/years. Weight-for-height in kilograms/meters.

[c] Stunting and wasting are based on height-for-age and weight-for-age. *Z*-scores calculated are based on CHS/CDC/WHO standards. "Moderate" refers to being more than one standard deviation below the means and "severe" more than two standard deviations below mean.

[d] Ownership of durables measures number of durables owned out of 15 asked.

[e] Job of mother/job of father: 1=no job; 2=temporary job; 3=permanent job.

Table 4: *(Men) Kenya. T-tests for differences in means in Kenya 1 data for those re-interviewed versus not re-interviewed* [a]

MEN:	Re-interviewed		Not re-interviewed		Difference	
Variables	Mean	Standard Deviation	Mean	Standard Deviation	Mean	t-test
Fertility-related outcome variables						
Currently using contraceptives	0.196	(0.017)		(0.031)	-0.033	(-0.95)
Ever used contraceptives	0.233	(0.018)	0.311	(0.052)	-0.077*	(-1.79)
Want no more children	0.208	(0.017)	0.237	(0.031)	-0.029	(-0.83)
Number of surviving children	4.76	(0.171)	3.94	(0.277)	0.817**	(2.46)
Family planning program variables						
Visited by community-based distribution agent	0.156	(0.015)	0.132	(0.025)	0.024	(0.78)
Heard family planning message on radio	0.931	(0.011)	0.968	(0.013)	-0.037*	(-1.86)
Heard about family planning at clinic	0.495	(0.021)	0.513	(0.036)	-0.018	(-0.42)
Discussed with others family planning lecture heard at clinic	0.679	(0.029)	0.691	(0.047)	-0.012	(-0.21)
Number of network partners in network for						
Family planning	3.7	(0.20)	4.0	(0.35)	-0.3	(-0.86)
Wealth flows	5.0	(0.21)	5.0	(0.36)	-0.04	
Reproductive health	–		–		–	(-0.10)
Knows secret contraceptive user	0.637	(0.069)	0.558	(0.095)	0.079	(0.60)
Control variables						
Age (years)	40.1	(0.52)	36.8	(0.78)	3.3**	(3.24)
Education						
No schooling	0.112	(0.013)	0.063	(0.018)	0.049*	(1.94)
Some primary schooling	0.577	(0.021)	0.537	(0.036)	0.040	(0.96)
Secondary schooling	0.298	(0.019)	0.379	(0.035)	-0.081**	(-2.06)
Language						
Luo only	0.796	(0.017)	0.805	(0.029)	-0.010	(-0.28)
English	0.443	(0.021)	0.532	(0.036)	-0.089**	(-2.11)
Swahili	0.655	(0.020)	0.726	(0.032)	-0.072*	(-1.82)
Lived						
outside of province	0.591	(0.021)	0.653	(0.035)	0.061	(1.49)
in Nairobi or Mombasa	0.336	(0.020)	0.400	(0.036)	-0.064	(-1.58)
Belongs to credit group	0.257	(0.019)	0.242	(0.031)	0.015	(0.40)
Belong to clan welfare society	0.868	(0.014)	0.905	(0.021)	-0.037	(-1.35)
Women sell on market	–		–		–	
Household characteristics						
Polygamous household	0.293	(0.019)	0.238	(0.031)	0.055	(1.45)
Self/Husband receives monthly salary	0.170	(0.016)	0.255	(0.032)	-0.085**	(-2.56)
Husband interviewed	–		–		–	
Household has radio	–		–			
House has metal roof	0.173	(0.016)	0.189	(0.029)	-0.016	(-0.51)
Sublocation of residence						
Gwassi	0.278	(0.019)	0.216	(0.030)	0.063*	(1.69)
Kawadhgone	0.230	(0.018)	0.237	(0.031)	-0.007	(-0.20)
Oyugis	0.259	(0.019)	0.300	(0.033)	-0.041	(-1.11)
Ugina	0.233	(0.018)	0.247	(0.032)	-0.014	(-0.39)

Table 4:   *(continued) (Women)*

WOMEN:	Re-interviewed		Not re-interviewed		Difference	
Variables	Mean	Standard Deviation	Mean	Standard Deviation	Mean	t-test
Fertility-related outcome variables						
Currently using contraceptives	0.126	(0.012)	0.103	(0.021)	0.024	(0.91)
Ever used contraceptives	0.238	(0.016)	0.196	(0.027)	0.042	(1.25)
Want no more children	0.351	(0.018)	0.220	(0.037)	0.132**	(3.59)
Number of surviving children	3.88	(0.089)	2.78	(0.138)	1.10**	(5.90)
Family planning program variables						
Visited by community-based distribution agent	0.163	(0.014)	0.113	(0.022)	0.050*	(1.75)
Heard family planning message on radio	0.870	(0.916)	0.916	(0.019)	-0.046*	(-1.79)
Heard about family planning at clinic	0.851	(0.013)	0.828	(0.027)	0.023	(0.80)
Discussed with others family planning lecture heard at clinic	0.629	(0.070)	0.661	(0.037)	-0.032	(-0.76)
Number of network partners in network for						
Family planning	2.9	(0.11)	3.1	(0.20)	-.18	(-0.78)
Wealth flows	2.8	(0.12)	2.4	(0.21)	0.38	(1.45)
Reproductive health	3.2	(0.16)	2.8	(0.23)	0.38	(1.19)
Knows secret contraceptive user	0.408	(0.02)	0.377	(0.03)	0.030	(0.77)
Control variables						
Age (years)	29.7	(0.332)	26.3	(0.488)	3.4**	(5.04)
Education						
No schooling	0.214	(0.015)	0.141	(0.024)	0.072*	(2.30)
Some primary schooling	0.669	(0.018)	0.668	(0.033)	0.001	(0.03)
Secondary schooling	0.117	(0.012)	0.190	(0.027)	-0.074**	(-2.75)
Language						
Luo only	0.422	(0.018)	0.327	(0.033)	0.095*	(2.46)
English	0.178	(0.014)	0.263	(0.031)	-0.086**	(-2.73)
Swahili	0.396	(0.018)	0.517	(0.035)	-0.121**	(-3.11)
Lived						
outside of province	0.370	(0.018)	0.371	(0.034)	-0.001	(-0.02)
in Nairobi or Mombasa	0.214	(0.015)	0.205	(0.028)	0.009	(0.29)
Belongs to credit group	0.351	(0.018)	0.288	(0.032)	0.064*	(1.70)
Belong to clan welfare society	0.747	(0.016)	0.644	(0.034)	0.103**	(2.93)
Women sell on market	0.464	(0.019)	0.444	(0.035)	0.020	(0.51)
Household characteristics						
Polygamous household	0.350	(0.018)	0.371	(0.034)	-0.021	(-0.56)
Self/Husband receives monthly salary	0.334	(0.019)	0.402	(0.037)	-0.068*	(-1.66)
Husband interviewed	0.765	(0.016)	0.752	(0.029)	0.013	(0.41)
Household has radio	0.492	(0.019)	0.546	(0.035)	-0.055	(-1.38)
House has metal roof	0.201	(0.015)	0.187	(0.027)	0.014	(0.45)
Sublocation of residence						
Gwassi	0.213	(0.015)	0.210	(0.029)	0.003	(0.08)
Kawadhgone	0.240	(0.015)	0.205	(0.028)	0.035	(1.06)
Oyugis	0.286	(0.017)	0.263	(0.031)	0.023	(0.63)
Ugina	0.261	(0.016)	0.322	(0.033)	-0.061*	(-1.72)

Note: * indicates significance at the 10 percent level, and ** at the 5 percent level.

[a] Values of two-sample t-test with unequal variances are given in parentheses in third and sixth columns.

Table 5: *South Africa. T-tests for differences in means in South Africa 1 data for those re-interviewed versus not re-interviewed [a]*

	Re-interviewed		Not re-interviewed		Difference	
	Mean	Standard Deviation	Mean	Standard Deviation	Means	t-test
Early child nutritional status and health outcome variables						
Height-for-age[b]	0.380	(0.009)	0.381	(0.017)	-0.001	(-0.08)
Weight-for-age[b]	5.400	(0.109)	5.328	(0.199)	0.072	(0.32)
Weight-for-height[b2]	14.80	(0.101)	14.69	(0.199)	0.111	(0.50)
Height-for-age z-score	-1.148	(0.073)	-1.282	(0.142)	0.134	(0.84)
Weight-for-age z-score	-0.616	(0.059)	-0.735	(0.108)	0.119	(0.97)
Weight-for-height z-score	0.167	(0.071)	0.078	(0.138)	0.090	(0.58)
Moderate stunting[c]	0.534	(0.019)	0.525	(0.036)	0.008	(0.21)
Severe stunting[c]	0.270	(0.017)	0.273	(0.032)	-0.002	(-0.07)
Moderate wasting[c]	0.388	(0.018)	0.444	(0.035)	-0.057	(-1.42)
Severe wasting[c]	0.187	(0.015)	0.172	(0.027)	0.016	(0.51)
Predetermined family background variables						
Age in months	37.12	(0.675)	37.08	(1.272)	0.044	(0.03)
Fraction male	0.499	(0.019)	0.495	(0.035)	0.004	(0.11)
Fraction African	0.910	(0.011)	0.859	(0.025)	0.051*	(1.89)
Household size	8.856	(0.147)	8.500	(0.296)	0.356	(1.08)
Total monthly expenditures	1483.4	(30.53)	1510.9	(63.63)	-27.46	(-0.39)
Per capita monthly expenditures	195.2	(5.612)	217.5	(13.17)	-22.33	(-1.56)
Total monthly income	1158.1	(45.26)	1391.0	(99.43)	-234**	(-2.13)
Per capita monthly income	156.3	(7.922)	216.6	(21.36)	-60.4**	(-2.65)
Household head age	51.77	(0.524)	52.64	(1.095)	-0.871	(-0.72)
Household head education	2.957	(0.125)	3.485	(0.255)	-0.528*	(-1.86)
Household head male	0.695	(0.017)	0.702	(0.033)	-0.007	(-0.18)
Own house	0.883	(0.012)	0.838	(0.026)	0.044	(1.53)
Number of rooms	4.951	(0.100)	5.318	(0.215)	-0.367	(-1.55)
Number of durables	3.149	(0.082)	3.556	(0.149)	-0.41**	(-2.39)
Urban	0.289	(0.017)	0.343	(0.034)	-0.054	(-1.44)
In former Natal	0.165	(0.014)	0.237	(0.030)	-0.07**	(-2.18)

Notes: * indicates significance at the 10 percent level, and ** at the 5 percent level.

[a] Values of two-sample t-test with unequal variances are given in parentheses in last column.

[b] Height-for-age in meter/years. Weight-for-age in kilogram/years. Weight-for-height in kilograms/meters.

[c] Stunting and wasting are based on height-for-age and weight-for-age. Z-scores calculated based on NCHS/CDC/WHO standards. "Moderate" refers to being more than one standard deviation below the means and "severe" more than two standard deviations below mean.

4.2 Probits for Probability of Attrition

We start with a parsimonious specification of probits for the probability of attrition in which only one outcome variable at a time is included; we then include all outcome variables plus predetermined family background variables (Table 6). The dependent variable in these probits is whether attrition occurred between the survey rounds (1=yes; 0=no); χ^2 tests for the significance of the overall relations are presented at the bottom of Table 6.

Bolivia: The χ^2 tests indicate that if only one of the outcome variables at a time is included in these probits, the probit is significant at the 5 percent level only for severe stunting, that is, a child who is severely stunted is more likely to be lost to follow-up. For moderate and severe low weight-for-age and the four test scores, the probits are significant at the 10 percent level, suggesting that poor childhood development is associated with higher probability of attrition. When all of the family background variables and all childhood development indicators are included in the analysis, however, among the childhood development indicators only moderate stunting is significantly nonzero, even at the 10 percent level, with a negative sign. That 1 in 11 of the childhood development indicators has a significant coefficient estimate at the 10 percent level in the multivariate analysis is what one would expect to occur by chance, even if none of the childhood development indicator coefficients were truly significant predictors of attrition. Moreover, the one childhood development outcome variable that has a significantly nonzero coefficient estimate in Table 6 in the multivariate analysis does *not* show significant differences in the comparison of means in Table 3.

The comparisons of means for childhood development outcomes between subsamples of those lost to follow-up and those who were re-interviewed, therefore, may be misleading regarding the extent of significant associations of these childhood development indicators with sample attrition once family background characteristics are controlled. The comparisons in Table 3 indicate that there is selective attrition with regard to childhood development indicators, with those children who are worse off in round 1 significantly more likely to be lost to follow-up. But the multivariate estimates present a different picture: they indicate that the extent of significant associations for the child development outcomes in probits for predicting attrition is about what would be expected by chance. Thus, conditional on controls for observed family background characteristics, attrition is *not* predicted by child development indicators for round 1. (Of course, there may be multicollinearity among the child development indicators that disguises their significance.)

If the predetermined family background variables in Bolivia 1 are included alone or with all of the early childhood development indicators, the probits are significantly nonzero at very high levels. Some family background variables are significantly (at least at the 10

Table 6: *Probits for predicting attrition between rounds 1 and 2 for Bolivian, Kenyan, and South African data [a]*

	Bolivia			Kenyan Men			Kenyan Women		South Africa		
Outcome variables	Outcome variables, one at a time	All outcome variables + pre-determined variables[b]	Outcome variables	Outcome variables, one at a time	All outcome variables + pre-determined variables[c]	Outcome variables, one at a time	All outcome variables + pre-determined variables[d]	Outcome variables	Outcome variables, one at a time	All outcome variables + pre-determined variables[e]
Height-for-age	-.0015 (-0.83)	-.0002 (-0.04)	Currently contracepting	0.118 (0.95)	-0.065 (0.34)	-0.134 (0.92)	0.004 (0.02)	Height-for-age	0.016 (0.09)	1.204 (1.30)
Weight-for-height	-.0015 (-0.99)	.0032 (0.80)	Ever used contraceptives	0.162* (1.67)	-0.103 (-0.70)	-0.142 (1.26)	-0.036 (0.28)	Weight-for-height	-0.009 (-0.45)	0.040 (1.02)
Weight-for-age	-.003* (-1.74)	-.0037 (-0.78)	Want no more children	0.099 (0.83)	0.245* (1.69)	-0.374** (3.60)	-0.010 (0.07)	Weight-for-age	-0.005 (-0.34)	-0.082 (-1.20)
Moderate wasting	.148* (1.78)	.1003 (0.70)	Number of surviving children	-0.033** (-2.46)	-0.017 (-0.78)	-0.139** (5.82)	-0.136** (3.73)	Moderate wasting	0.136 (1.25)	0.297* (1.67)
Severe wasting	.191 (1.35)	.1353 (0.70)	Number of family planning network partners	-0.009 (-0.85)	0.003 (0.22)	0.012 (0.78)	-0.010 (0.56)	Severe wasting	-0.062 (-0.52)	-0.144 (-0.95)
Moderate stunting	-.0315 (-0.38)	-.291* (-1.93)						Moderate stunting	-0.019 (-0.21)	-0.036 (-0.33)
Severe stunting	.2110** (2.41)	.2066 (1.51)						Severe stunting	0.007 (0.06)	0.005 (0.03)
Bulk motor ability	-.009 (-1.64)	.0123 (0.59)								
Fine motor ability	-.009 (-1.63)	-.0073 (-0.35)								
Language-audition	-.010* (-1.84)	-.0059 (-0.27)								
Personal-social	-.008 (-1.64)	-.0014 (-0.07)								
Constant		.75* (1.72)			-0.239 (-0.70)		-0.097 (0.29)			-0.989 (-0.72)
χ^2test [prob > χ^2]	f	300.22 [0.001]		g	25.13 [0.068]	h	54.49 [0.001]		i	6.67 [0.464]

Table 6: (notes)

Note: * indicates significance at the 10 percent level, and ** indicates significance at the 5 percent level.

a Values of z-tests are in parentheses beneath point estimates. P-values of Chi-square tests are in brackets.

b Predetermined variables for Bolivian households that are: (a) significant at 5 percent level (with sign in parentheses)—father's age(+); Quechua only (+); ownership of house (-); number of durables owned (-); Oruro (-), Postosi (-), Santa Cruz (-) relative to La Paz; mother's job permanent relative to no job (-); (b) significant at the 10 percent level – father's schooling (-), number of rooms in the house (+), number of siblings of child (-); father's job temporary relative to no job (-); (c) not significant even at the 10 percent level – mother's age, mother's schooling, Amarya only, El Alto, Cochabamba, Tarija relative to La Paz; father's job permanent relative to no job; mother's job temporary relative to no job; household income.

c Predetermined variables for Kenyan men that are (a) significant at the 5 percent level (with sign in parentheses)—men's age; (b) not significant even at the 10 percent level – primary schooling; secondary schooling; Luo only; English; lived in Nairobi or Mombasa; polygamous household; earns a monthly salary; sublocation of residence.

d Predetermined variables for Kenyan women that are: (a) significant at the 5 percent level (with sign in parentheses)—husband interviewed (-); (b) significant at the 10 percent level—resided in Oyugnis relative to Ugina (-) (c) not significant even at the 10 percent level—primary schooling; secondary schooling; Luo only; English; lived in Nairobi or Mombasa; polygamous household; household has radio; household has metal roof; other sublocation of residence.

e Predetermined variables for South African households that are (a) significant at the 5 percent level (with sign in parentheses)—age of household head(+); (b) significant at the 10 percent level—none; (c) not significant even at the 10 percent level—male child; African household; household size; ln total monthly expenditures; household head schooling; male household head; own the house; number of rooms; number of durables; urban; former Natal.

f For Bolivian data, Probability > χ2 (a) at the 5 percent level—severe stunting; (b) at the 10 percent level—weight-for-age, moderate wasting, language-auditory.

g For Kenyan men, Probability > χ2 (a) at the 5 percent level—number of surviving children; (b) at the 10 percent level—ever-used contraceptives.

h For Kenyan women, Probability > χ2 (a) at the 5 percent level—want no more children, number of surviving children.

i For South African data, Probability > χ^2 (a) at the 5 percent level—none; (b) at the 10 percent level—none.

percent level) associated with higher probability of attrition: older and less-schooled fathers, speaking mainly Quechua in the household, not owning the home, having more rooms in the house, having fewer siblings, having fewer durables, father having permanent or no (rather than a temporary) job, and mother having no or a temporary (rather than a permanent) job, with some significant differences also among the urban areas included in the program. The majority of these significant coefficient estimates are consistent with what might be predicted from the significant differences in the means in Table 3, reinforcing the observation that attrition tends to be selectively greater among children from worse-off family backgrounds.

But some of these significant coefficient estimates are opposite in sign from what might be expected from the comparisons of the means in Table 3, suggesting the opposite relation to attrition if there are multivariate controls for standard background variables other than what appear in the comparisons of means. Specifically, the comparisons in Table 3 suggest that attrition is significantly more likely if fathers are younger, the house has fewer rooms, and there are fewer siblings, but all three of these signs are reversed with significant coefficient estimates in the multivariate analyses of Table 6. Moreover, two variables that are not significantly different for the two subsamples in Table 3 have significant coefficient estimates in Table 6, i.e., father's schooling and father having a temporary job, both of which are estimated to significantly reduce attrition probabilities in Table 6. Finally, both mother's age and household income have means that are significantly different between the subsamples in the univariate comparisons in Table 3, but do not have coefficient estimates that are significantly nonzero, even at the 10 percent level, once there is control for other family background characteristics in Table 6.

Thus, exactly which family background characteristics predict attrition with multivariate controls and what the directions of those effects are cannot be inferred simply by examining the significance of means in univariate comparisons between the subsamples. While the patterns in Tables 3 and 6 suggest that worse-off family background is associated with greater attrition, the multivariate estimates are less supportive of this conclusion.

Kenya: Since there are gender differences in the probit estimates of the probability of attrition, we report separately for men and women (Table 6). For men, we find that when the five outcomes are included singly, only the number of surviving children is significantly related to attrition at the 5 percent level; one other – ever-used family planning – is significantly related to attrition at the 10 percent level. If other right-side variables are included, among the five fertility related outcomes none is significantly nonzero at the 5 percent level, and only not wanting more children is significantly related to attrition at the 10 percent level. A χ^2 test for the joint significance of these five variables rejects such significance (p=0.52). Among the control variables only age is significant, but not schooling, language, household characteristics, past residence in Nairobi or Mombasa, or

current sublocation of residence. A χ^2 test for the joint significance of all the right-side variables rejects such significance at the 5 percent level (p=0.068).

For women, we find that two of the lagged outcome variables, wanting no more children and the number of surviving children, are individually significant (and negative). When all the lagged outcome variables and the predetermined variables are included, only the latter (number of surviving children) remains significant. However, in contrast to the results for men, χ^2 tests for the joint significance of the five fertility related outcome variables and for the entire set of right-side variables indicate significance (p < 0.0001 in both cases).

Thus, for the Kenyan data, there is no significant association between attrition, most of the outcome variables, and most of the major control variables. However, gender does matter in these multivariate analyses: there is a significant negative association between attrition and number of surviving children for women but not for men.

South Africa: Probit estimates for the probability of attrition reveal little evidence that the outcome variables are associated with attrition of pre-school children, paralleling the results of the mean comparisons presented in Section 4.1. When only one outcome variable at a time is included, none is significant at conventional levels. When the set of outcome variables are included at the same time, all but moderate wasting are insignificant and a joint χ^2 test indicates that the set of all outcome variables together is insignificant. Moreover, the overall relation is insignificant – this set of background characteristics and outcome variables does a very poor job predicting attrition in the sample. Thus, for the South African data, there is no significant association between attrition of pre-school children, most of the outcome variables, and most of the major control variables.

4.3 Do Those Lost to Follow-up have Different Coefficient Estimates than Those Re-interviewed?

Our aim here is to determine whether those who subsequently leave the sample differ in their initial behavioral relationships. We conduct the BGLW tests, in which the value of an outcome variable at the initial wave of the survey is regressed on predetermined variables for the initial survey wave and on subsequent attrition. In short, the test is whether the coefficients of the predetermined variables and the constant differ for those respondents who are subsequently lost to follow- up versus those who are re-interviewed. Tables 7, 8, and 9 present these multivariate regression and probit estimates for the same outcome variables considered above, with the same family background variables as controls. The first part of each table gives the coefficient estimates for the family background variables for the subsample of those who were re-interviewed. At the bottom of each table are the F or χ^2 tests (for ordinary least squares regression or probit, respectively) for whether there

are significant differences between the two subsamples that test for equality of (i) all of the slope coefficients and the constant and (ii) all of the slope coefficients (but not the constant).

Bolivia: F tests indicate that all of the eleven estimated equations for childhood development indicators are statistically significant with a p-value of $p < 0.0001$ (Table 7). These estimates indicate a number of associations that are consistent with widely held perceptions about child development. For example, household income is significantly positively associated with height-for-age and significantly negatively associated with severe stunting; mother's schooling is significantly positively associated with height-for-age and weight-for-age, though significantly negatively associated with gross motor ability; and ownership of consumer durables is significantly positively associated with height-for-age, gross motor ability, fine motor ability, language-audition, and personal-social test scores, but significantly negatively associated with severe wasting.

There are, however, no significant differences at the 5 percent level (Note 15) between the set of coefficients for the subsample of those lost to follow-up versus the subsample of those re-interviewed for over half of the indicators of child development: height-for-age, moderate stunting, gross motor ability tests, fine motor ability tests, language-audition tests, and personal-social tests. The second set of tests, further, indicates that there are no significant differences at the 10 percent level for severe stunting. These estimates for the anthropometric indicators related to stunting and for the four cognitive development test scores, therefore, suggest that the coefficient estimates of standard family background variables are *not* significantly affected by sample attrition.

The results differ sharply, however, for the anthropometric indicators related to wasting. Both tests for these four child outcome variables indicate that the coefficient estimates for observed family background variables do differ significantly at the 5 percent level (and for all but weight-for-age at the 1 percent level) between the two subsamples. For these outcomes, therefore, it is important to control for the attrition in the analysis, e.g., as with the matching methods used in Behrman, Cheng and Todd (2001).

Kenya: We conduct BGLW tests with Kenya 1 contraceptive use (ever or current), want no more children, number of surviving children, and family planning network size as the dependent variables (Table 8). The right-side variables again include a fairly standard set of control variables, i.e., age, schooling, wealth indicators, language indicators, and location of residence. Tests for the significance of the differences in the slope coefficients in all cases for both men and women fail to reject equality of all the coefficients between the subsamples of those lost to follow-up and those re-interviewed. Tests for the joint significance of the differences in the slope coefficients and intercepts in all cases fail to reject equality of all the coefficients and of an additive variable for attrition (with the exception at the 5 percent level of number of surviving children and at the 10 percent level

for currently using contraceptives, both only for women and in both of which cases the constant differs between the subsamples, but not the slope coefficient estimates).

Thus there is no significant effects on the slope coefficients of attrition for either men or women, and but limited evidence of a significant effect on the constants for women.

South Africa: The evidence for South Africa presented earlier in Sections 4.1 and 4.2 suggests that attrition bias resulting from selection on observables is not present. The BGLW tests examined in this section largely confirm this, although there are some exceptions.

For the first three anthropometric outcomes shown in Table 9, the attrition interactions are not jointly significant with or without the attrition dummy variable. In the remaining columns that present the stunting and wasting probits, the attrition interaction terms are significant only in the case of moderate stunting, indicating the possibility of attrition bias in this relationship. On the other hand, attrition does not appear to have any association with severe stunting or moderate and severe wasting.

As described in Section 3, one important difference in the South African sample relative to the others is that, when possible, households that had moved were followed. These households are included in the analysis presented above. What would happen if they were excluded? Re-estimating the equations in Table 9 categorizing those who had moved but were interviewed as if they had been lost to follow-up and not re-interviewed leads to a somewhat stronger, but still fairly weak, rejection of the null hypothesis that there are no differences in coefficients across the two groups (results not shown). In every case the p-values for either the F or χ^2 tests on the attrition interactions decline; for height-for-age, weight-for-age, and moderate wasting the effect of attrition on the constant becomes significant at the 10 percent level. It appears that the investment made in following movers had some payoff in terms of reduced attrition bias for this set of relationships, though these alternative estimates still do not indicate very high probabilities of attrition bias and where it exists, it is concentrated in a shift in the constant term.

Table 7a: *Bolivia. Testing impact of attrition between Bolivia 1 and Bolivia 2 on coefficient estimates of family background variables in early childhood development anthropometric outcomes[a]*

Right-side variables	Ordinary Least Squares Regressions for				Probits for			
	Height for age	Weight for age	Weight for height	Moderate Stunting	Severe Stunting	Moderate Wasting	Severe Wasting
Predetermined Family Background Variables							
Mother's age	-0.0369	0.162	0.214	-0.00933	-.00363	-0.00352	0.0142
	(-0.31)	(1.13)	(1.46)	(-0.79)	(-0.27)	(-0.29)	(0.67)
Father's age	0.222**	0.130	-0.072	-0.00558	-0.0165	-.0209**	-0.0186
	(2.29)	(1.13)	(-0.61)	(-0.58)	(-1.50)	(-2.08)	(-1.06)
Mother's schooling	0.998**	1.51**	0.611	—	—	—	—
	(2.40)	(3.05)	(1.20)				
Father's schooling	-0.143	-0.407	-0.534	—	—	—	-0.106
	(-0.34)	(-0.82)	(-1.05)				(-1.37)
Quechua mainly	-3.58	-7.23	-1.05	16.4**	-0.667	17.3**	—
	(-0.23)	(-0.40)	(-0.06)	(21.42)	(-0.46)	(25.26)	
Amarya mainly	-0.010	-3.19	-7.47	-0.755	0.476	0.313	—
	(-0.00)	(-0.35)	(-0.79)	(-1.00)	(0.65)	(0.43)	
Ownership of house	-1.37	-1.07	0.075	0.0537	0.0183	-0.0225	—
	(-1.20)	(-0.79)	(0.05)	(0.46)	(0.15)	(-0.20)	
Number of rooms in the house	1.48**	1.15	0.108	-0.0523	-0.0591	-0.0127	-0.0269
	(2.44)	(1.59)	(0.15)	(-0.86)	(-0.83)	(-0.21)	(-0.23)
Number of siblings	-1.76**	-1.50**	0.133	0.182**	0.242**	0.104**	—
	(-5.08)	(-3.63)	(0.31)	(4.99)	(6.42)	(3.00)	
Ownership of durables	0.946**	0.535	-0.246	—	—	—	-0.172**
	(3.28)	(1.56)	(-0.70)				(-3.13)
El Alto	0.036	-0.135	2.149	.262*	0.343**	-0.0610	-0.150
	(0.03)	(-0.08)	(1.182)	(1.70)	(2.22)	(-0.42)	(-0.54)
Cochabamba	4.63**	-2.17	-6.01**	—	—	0.130	—
	(2.94)	(-1.16)	(-3.12)			(0.84)	
Oruro	-4.43**	-6.89**	1.12	0.526**	0.551**	0.509**	0.676**
	(-2.10)	(-2.75)	(0.44)	(2.29)	(2.56)	(2.53)	(2.10)
Potosi	-0.869	-10.0**	-11.93**	0.229	0.481**	0.936**	—
	(-0.43)	(-4.16)	(-4.83)	(1.08)	(2.34)	(4.78)	
Tarija	6.65**	14.35**	12.4**	-0.189	-0.0944	-0.723**	—
	(3.18)	(5.76)	(4.83)	(-0.91)	(-0.41)	(-3.10)	
Santa Cruz	9.65**	5.02**	-2.27	-0.748**	-0.673**	-0.346**	-0.372
	(6.28)	(2.74)	(-1.21)	(-4.92)	(-3.67)	(-2.21)	(-1.26)

Table 7a: *(continued)*

Right-side variables	Ordinary Least Squares Regressions for				Probits for		
	Height for age	Weight for age	Weight for height	Moderate Stunting	Severe Stunting	Moderate Wasting	Severe Wasting
Job of father is temporary	-4.77* (-1.79)	-7.29** (-2.30)	-3.85 (-1.18)	0.411 (1.57)	0.6766* (2.06)	0.372 (1.35)	—
Job of father is permanent	-4.38* (-1.73)	-6.38** (-2.12)	-2.88 (-0.93)	0.393 (1.59)	0.679** (2.14)	0.282 (1.07)	0.0729 (0.16)
Job of mother is temporary	-4.80** (-2.84)	-3.53* (-1.75)	2.63 (1.27)	0.544** (3.04)	0.692** (3.90)	0.268* (1.61)	0.0967 (0.33)
Job of mother is permanent	-3.23** (-2.91)	-1.92 (-1.46)	2.37* (1.75)	0.250** (2.26)	0.390** (3.07)	0.226** (2.01)	0.0356 (0.18)
Household income	.00121* (1.62)	.000558 (0.63)	-.000538 (-0.59)	-0.000065 (-0.86)	-0.000164* (-1.64)	-0.0000262 (-0.33)	-0.0000376 (-0.25)
Constant	10.28** (2.51)	27.19** (5.58)	57.91** (11.58)	0.845** (2.07)	-0.901* (-1.87)	-0.00232 (-0.01)	-1.39* (-1.91)
F test for overall relation [probability > F test]	7.11** [0.0001]	5.58 ** [0.0001]	4.02** [0.0001]	257.80** [0.0001]	278.38** [0.0001]	179.06** [0.0001]	98.91** [0.0001]
F Tests for attrition [probability > F]							
1. Joint effect of attrition on constant and all estimates	1.32 [0.1428]	1.88** [0.0070]	1.58** [0.0385]	22.68 [0.3614]	35.34* [0.0357]	44.86** [0.0018]	261.66** [0.0001]
2. Joint effect of attrition on all coefficient estimates but not on constant	1.37 [0.1169]	1.90** [0.0068]	1.63** [0.0315]	22.49 [0.3147]	29.18 [0.1097]	42.17** [0.0026]	253.89** [0.0001]

Note:

* indicates significance at the 10 percent level, and ** indicates significance at the 5 percent level. P-values of tests are in brackets.

[a] Values of t-tests (for regressions) and *z*-tests (for probits) are in parentheses beneath point estimates.

Table 7b: *Bolivia. Multivariate ordinary least squares regressions for testing impact of attrition between Bolivia 1 and Bolivia 2 on coefficient estimates of family background variables in child test scores[a]*

Right-side variables	Gross motor ability	Fine motor ability	Language-auditory	Personal-social
Predetermined Family Background Variables				
Mother's age	0.204**	0.189**	0.203**	0.199**
	(4.84)	(4.80)	(4.96)	(4.57)
Father's age	-0.00767	0.00268	0.0118	0.00547
	(-0.23)	(0.08)	(0.36)	(0.16)
Mother's schooling	-0.257*	-0.127	-0.0290	-0.167
	(-1.75)	(-0.93)	(-0.20)	(-1.10)
Father's schooling	0.236*	0.219	0.159	0.209
	(1.61)	(1.60)	(1.12)	(1.38)
Quechua mainly	2.85	2.88	3.32	4.28
	(0.53)	(0.57)	(0.63)	(0.77)
Amarya mainly	-4.01	-3.05	-3.091	-2.91
	(-1.47)	(-1.19)	(-1.17)	(-1.03)
Ownership of house	-0.167	0.137	-0.123	—
	(-0.41)	(0.36)	(-0.31)	
Number of rooms in the house	-0.0260	0.0373	-0.0751	0.0433
	(-0.12)	(0.19)	(-0.36)	(0.20)
Number of siblings	-0.0370	-0.139	-0.00220	-0.103
	(-0.30)	(-1.21)	(-0.02)	(-0.81)
Ownership of durables	0.335**	0.278*8	0.395**	0.403**
	(3.30)	(2.92)	(4.00)	(3.84)
El Alto	1.70**	1.49**	1.87**	1.84**
	(3.26)	(3.07)	(3.71)	(3.43)
Cochabamba	0.569	-0.254	0.156	0.675
	(1.03)	(-0.49)	(0.29)	(1.18)
Oruro	.537	-0.337	0.761	0.401
	(0.72)	(-0.49)	(1.06)	(0.52)
Potosi	-1.08	-1.23*	-0.720	-1.07
	(-1.51)	(-1.85)	(-1.04)	(-1.45)
Tarija	4.01**	2.64**	3.31**	3.68**
	(5.43)	(3.83)	(4.63)	(4.83)
Santa Cruz	2.05**	1.09**	1.63**	—
	(3.79)	(2.16)	(3.10)	

Table 7b: *(continued)*

Right-side variables	Gross motor ability	Fine motor ability	Language-auditory	Personal-social
Predetermined Family Background Variables				
Job of father is temporary	—	-1.79*	-1.77*	-1.69*
		(-2.05)	(-1.95)	(-1.75)
Job of father is permanent	-2.35**	-2.03**	-2.09**	-2.02**
	(-2.64)	(-2.44)	(-2.42)	(-2.20)
Job of mother is temporary	2.20**	1.92**	---	2.17**
	(3.69)	(3.45)		(3.53)
Job of mother is permanent	0.948**	0.900**	0.844**	1.06**
	(2.43)	(2.45)	(2.22)	(2.63)
Household income	.000068	.0000878	-0.0000282	-0.0000404
	(0.26)	(0.36)	(-0.11)	(-0.15)
Constant	13.4**	12.47 **	10.28**	11.4**
	(9.28)	(9.25)	(7.35)	(7.62)
F-test for overall relation [probability > F-test]	5.38** [0.0001]	5.21** [0.0001]	5.80** [0.0001]	5.39** [0.0001]
F-Tests for Attrition [probability > F]				
1. joint effect of attrition on all estimates, including constant	1.31 [0.1461]	1.45* [0.0772]	1.34 [0.1277]	1.38 [0.1055]
2. joint effect of attrition on all coefficients but not on constant	1.37 [0.1160]	1.51* [0.0594]	1.40 [0.1013]	1.44* [0.0824]

Note:

* indicates significance at the 10 percent level, and ** indicates significance at the 5 percent level. P-values of tests are in brackets.

[a] Values of t-tests are in parentheses beneath point estimates.

Table 8: *(Men) Kenya. Multivariate probits/regressions for testing impact of attrition for men and women between Kenya 1 and Kenya 2 on key fertility-related outcome variables [a]*

Right-side variables (MEN)	Probits			OLS Regressions	
	Currently using contraceptives	Ever used contraceptives	Want no more children	Number of surviving children	Family planning social network size
Control variables					
Age (years)	0.004 (0.74)	0.009 (1.62)	0.013** (8.58)	0.200** (20.26)	0.015 (0.86)
Education (relative to no schooling)					
Primary schooling	0.075 (0.36)	-0.048 (0.26)	0.133 (0.69)	0.955** (2.85)	1.202** (2.08)
Secondary schooling	0.310 (1.22)	0.122 (0.55)	0.197 (0.81)	0.736* (1.77)	2.247** (3.12)
Language					
Luo only	0.372* (1.87)	0.368** (2.37)	0.142 (0.89)	-0.180 (0.66)	0.815* (1.74)
English	-0.037 (0.24)	-0.048 (0.33)	0.074 (0.46)	0.325 (1.20)	0.243 (0.52)
Lived in Nairobi or Mombasa	0.130 (1.12)	0.221** (2.02)	0.324** (2.74)	0.086 (0.41)	0.258 (0.71)
Women sell in market	—	—	—	—	—
Household characteristics					
Polygamous household	0.091 (0.65)	-0.025 (0.19)	-0.296** (2.10)	2.386** (9.69)	0.017 (0.04)
Earns a monthly salary	0.058 (0.38)	0.302** (2.16)	0.251 (1.63)	0.312 (1.13)	0.953** (2.00)
Husband interviewed	–	–	–	–	--
Household has radio	–	–	–	–	--
Household has metal roof	–	–	–	–	
Sublocation of residence (relative to Ugina)					
Gwassi	-0.639** (3.42)	-0.571** (3.50)	-0.630** (3.42)	-0.032 (0.11)	-0.323 (0.66)
Kawadhgone	0.145 (0.88)	0.015 (0.09)	0.153 (0.93)	0.165 (0.57)	-0.182 (0.36)
Oyugis	0.256 (1.62)	0.239* (1.67)	0.328** (2.10)	0.229 (0.82)	-0.392 (0.81)
Constant	-1.53** (4.38)	-1.43** (4.67)	-3.34** (9.31)	-4.96** (8.94)	0.970 (1.02)
χ^2 test for overall relation [probability > χ^2]	48.87** [0.0001]	58.21** [0.0001]	134.25** [0.0001]		
R-squared / F-test [probability > F]				0.560 / 82.81** [0.0001]	0.057 / 3.98** [0.0005]
Tests for Attrition					
Effect of attrition on constant	0.027 (0.21)	0.046 (0.38)	0.150 (1.13)	-0.065 (0.29)	0.166 (0.42)
χ^2 test for joint effect of attrition on constant and all coefficient estimates [probability > χ^2] (F tests for regressions)	12.11 [0.437]	11.27 [0.506]	16.79 [0.158]	1.11 [0.352]	0.71 [0.725]
χ^2 test for joint effect of attrition on all coefficient estimates but not on constant [probability > χ^2] (F-tests for regressions)	11.90 [0.371]	11.04 [0.440]	15.27 [0.171]	1.20 [0.284]	0.67 [0.781]

Table 8:   *(Women) (Continued)*

Right-side variables (WOMEN)	Probits			OLS Regressions	
	Currently using contraceptives	Ever used contraceptives	Want no more children	Number of surviving children	Family planning social network size
Control variables					
Age (years)	0.014** (2.03)	0.023** (3.68)	0.079** (11.80)	0.161** (20.82)	0.025** (1.97)
Education (relative to no schooling)					
Primary schooling	0.122 (0.72)	0.094 (0.66)	-0.004 (0.03)	-0.440** (2.66)	0.957** (3.41)
Secondary schooling	0.125 (0.47)	0.279 (1.23)	-0.107 (0.46)	-0.447 (1.60)	1.786* (3.83)
Language					
Luo only	-0.268* (1.86)	-0.236* (1.95)	-0.228* (1.88)	-0.142 (1.00)	-0.395* (1.68)
English	0.264 (1.41)	0.265 (1.59)	-0.002 (0.01)	-0.334 (1.59)	0.125 (0.36)
Lived in Nairobi or Mombasa	0.311** (2.33)	0.356** (3.05)	0.240** (2.01)	0.144 (0.97)	-0.066 (0.26)
Women sell in market	0.254** (2.02)	0.147 (1.34)	-0.119 (1.07)	0.032 (0.24)	0.180 (0.83)
Household characteristics					
Polygamous household	-0.161 (1.28)	-0.104 (0.97)	0.187* (1.79)	-0.201 (1.57)	-0.089 (0.42)
Earns a monthly salary	—	—	—	—	—
Husband interviewed	0.211 (1.51)	-0.108 (0.94)	-0.113 (0.99)	-0.147 (1.05)	0.101 (0.44)
Household has radio	-0.019 (0.16)	-0.005 (0.05)	0.046 (0.44)	-0.106 (0.85)	0.270 (1.31)
Household has metal roof	0.003 (0.019)	0.253* (2.00)	0.173 (1.39)	0.810** (5.15)	0.142 (0.53)
Sublocation of residence (relative to Ugina)					
Gwassi	-0.441** (2.37)	-0.645** (4.10)	0.169 (1.13)	0.357* (2.03)	-0.668* (2.29)
Kawadhgone	-0.170 (0.99)	-0.260* (1.79)	0.130 (0.85)	0.240 (1.34)	0.496* (1.68)
Oyugis	0.013 (0.08)	-0.179 (1.26)	0.437** (2.93)	0.218 (1.23)	1.537** (5.22)
Constant	-1.85** (5.50)	-1.34** (4.71)	-3.03** (10.01)	-0.90** (2.57)	1.87** (3.23)
χ^2 test for overall relation [probability > χ^2]	44.22** [0.0001]	86.05** [0.0001]	234.12** [0.0001]		
R-squared / F-test [probability > F]				0.469 / 50.36** [0.0001]	0.082 / 5.48** [0.0001]
Tests for Attrition					
Effect of attrition on constant	0.126* (1.90)	-0.162 (1.31)	-0.189 (1.50)	-0.549** (3.77)	0.057 (0.24)
χ^2 test for joint effect of attrition on constant and all coefficient estimates [probability > χ^2] (F tests for regressions)	10.85 [0.763]	12.60 [0.633]	10.68 [0.775]	2.08** [0.009]	0.82 [0.657]
χ^2 test for joint effect of attrition on all coefficient estimates but not on constant [probability > χ^2] (F-tests for regressions)	10.74 [0.706]	11.58 [0.640]	9.20 [0.818]	1.05 [0.397]	0.87 [0.588]

Notes: * indicates significance at the 10 percent level, and ** at the 5 percent level.

[a] Absolute values of t-tests (for regressions) and z-tests (for probits) are in parentheses beneath point estimates:

Table 9: *South Africa. Multivariate regressions/probits for testing impact of attrition between South Africa 1 and South Africa 2 on child nutritional status and health* [a]

	Height-for-age	Weight-for-age	Weight-for-height	Moderate stunting	Severe stunting	Moderate wasting	Severe wasting
Control variables							
Respondent male	0.017	0.213	-0.032	0.094	0.118	0.156	0.133
	(0.94)	(0.90)	(0.15)	(0.99)	(1.12)	(1.44)	(1.14)
Respondent African	0.022	0.675	1.037**	0.038	0.082	-1.022**	0.107
	(0.55)	(1.39)	(2.57)	(0.13)	(0.21)	(3.43)	(0.31)
Household size	0.002	-0.020	-0.080**	0.009	-0.022	0.022	0.014
	(0.42)	(0.34)	(2.14)	(0.51)	(0.86)	(1.06)	(0.61)
Log total monthly expenditures	-0.001	0.093	0.276	-0.151	-0.191	-0.224	-0.009
	(0.03)	(0.35)	(1.18)	(1.25)	(1.24)	(1.45)	(0.06)
Household head age	-0.000	0.004	0.005	-0.004	0.005	0.001	0.003
	(0.01)	(0.32)	(0.47)	(0.85)	(0.92)	(0.28)	(0.60)
Household head schooling	-0.003	-0.058	-0.042	-0.019	0.009	0.014	0.013
	(0.78)	(1.20)	(1.00)	(0.90)	(0.35)	(0.66)	(0.48)
Household head male	-0.015	-0.312	-0.188	-0.025	0.012	0.147	0.242*
	(0.85)	(1.36)	(0.75)	(0.22)	(0.09)	(1.22)	(1.90)
Own house	-0.016	-0.257	-0.833**	0.103	0.454**	0.634**	0.703**
	(0.56)	(0.71)	(2.96)	(0.64)	(2.03)	(3.76)	(3.02)
Number of rooms	0.000	0.044	0.090*	-0.011	0.024	-0.041	-0.051**
	(0.04)	(1.03)	(1.73)	(0.54)	(0.98)	(1.56)	(2.14)
Number of durables	0.001	0.052	0.089	-0.044	-0.062	-0.076*	-0.064
	(0.15)	(0.69)	(1.24)	(1.15)	(1.34)	(1.89)	(1.38)
Urban	-0.007	-0.307	-0.376	-0.105	0.020	0.224	0.375**
	(0.35)	(1.29)	(0.94)	(0.60)	(0.09)	(1.28)	(1.88)
Former Natal	0.038	0.593	0.284	-0.281	-0.317	-0.524*	-0.343
	(1.14)	(1.43)	(0.96)	(1.64)	(0.99)	(1.90)	(1.15)
Constant	0.339**	4.207**	12.7**	1.440	0.221	1.651	-1.767
	(2.46)	(2.26)	(8.45)	(1.60)	(0.19)	(1.55)	(1.61)
F-test overall (Cols 1-3)	1.61*	2.25**	1.52*	106.8**	75.4**	76.1**	51.9**
χ^2 test overall (Columns 4-7) [p-value]	[0.065]	[0.005]	[0.092]	[0.001]	[0.001]	[0.001]	[0.001]
Tests for Attrition							
Effect of attrition on constant	0.359	4.212	2.783	-4.858**	-2.772	-2.469	0.660
	(1.25)	(1.24)	(0.47)	(2.19)	(1.11)	(1.13)	(0.28)
	[0.215]	[0.220]	[0.637]	[0.028]	[0.268]	[0.257]	[0.779]
Test for joint effect of attrition on constant and all estimates [p-value]	1.11 [0.364]	1.13 [0.353]	0.88 [0.576]	24.8** [0.024]	15.1 [0.301]	9.2 [0.760]	5.8 [0.954]
Test for joint effect of attrition on all estimates but constant [p-value]	1.18 [0.313]	1.21 [0.294]	0.91 [0.541]	24.8** [0.016]	15.1 [0.238]	5.4 [0.945]	5.6 [0.935]

Notes: * indicates significance at the 10 percent level, and ** at the 5 percent level. P-values of tests are in brackets.

Columns 1–3 are ordinary least squares and columns 4–7 are probit estimation. All are estimated allowing for clustering at community level and with robust standard errors to account for multiple observations on the same households within communities.

[a] Absolute values of t-tests (for regressions) and z-tests (for probits) are in parentheses below point estimates.

5. Conclusions

Our conclusions are similar in some respects to those of Fitzgerald, Gottschalk, and Moffitt (1998) for the Panel Study of Income Dynamics in the United States that is summarized in Section 2 but differ in other respects:

(a) The means for a number of critical child development outcome and family background variables do differ significantly between the subsample of those lost to follow-up between two rounds of a survey and those who were re-interviewed. For the Bolivian PIDI data, there is a definite tendency for those lost to follow-up to have poorer child development outcomes and family background than those who were re-interviewed. In the poor urban communities on which PIDI concentrates, it appears that worst-off households are most mobile and thus most difficult to follow over time. This is similar to the U.S. results. It contrasts, however, with the Kenyan rural data and the South African rural and urban data, where households and individuals with better backgrounds, (e.g., more schooling, more likely to speak English), are most mobile and thus hardest to follow over time. For the Kenyan data, this may be the case because better-off individuals tend to migrate from the poor rural sample areas to urban areas. For the South African data, however, this result is for both rural and urban areas, so it does not only reflect selective migration from rural to urban areas by those who are better off, but also perhaps selection for migration within urban areas.

(b) Neither family background variables nor outcome variables measured in the first of two surveys reliably predict attrition in multivariate probits. Some of the Bolivia 1 family background variables, but not the Bolivia 1 child outcome variables, are significant predictors of attrition. The result for the child outcome variables is similar to that for the outcome variables in the Kenyan case. But the significance of a number of background variables in predicting attrition in the Bolivian data, while similar to the U.S. results, again contrasts with the limited significance of such background variables in predicting attrition in the Kenyan and South African data. There are some gender differences in the Kenyan data, with attrition for women being more associated with their observed characteristics than is attrition for men.

(c) Attrition does not generally significantly affect the estimates of the association between family background variables and outcome variables. The coefficient estimates for standard family background variables in regressions and probit equations for the majority of the Bolivian child development outcome variables, including all of those related to stunting and to the test scores for gross and fine motor ability, language/auditory and personal/social interactions, are not affected significantly by attrition. The coefficients on standard variables in equations with the major outcome and family planning social network variables in the Kenyan data also are unaffected by attrition and, in contrast to the Fitzgerald, Gottschalk, and Moffitt (1998) study, the constants also do not differ (with the

possible exceptions of number of surviving children and of currently using contraceptives for which cases the constants differ at the 10 percent level for women). For five of the six child anthropometric measures in the South African data, moreover, there are no significant effects of attrition on the coefficient estimates of the standard variables nor, again, of the constants. Therefore, attrition apparently is not a general problem for obtaining consistent estimates of the coefficients of interest for most of the child development outcomes in the Bolivian data, for the fertility/social network outcomes in the Kenyan data, and for some of the anthropometric indicators in the South African data. These results are very similar to the results for the outcome measures for similar analyses with longitudinal U.S. data and suggest that despite suggestions of systematic attrition from univariate comparisons between those lost to follow-up and those re-interviewed, multivariate estimates of behavioral relations of interest may not be biased due to attrition.

It should be noted that for some outcomes the results differ strikingly and suggest that attrition bias will sometimes be a problem in multivariate estimates of behavioral relations that do not control for attrition. Among the particular outcomes that we consider in all three samples, there are significant interactions of attrition with the sets of standard variables that we consider in 5 out of 28, or 18 percent, of the cases, higher than the 5 percent that would be expected by chance at the 5 percent significance level. Attrition selection bias appears to be model specific: changing outcome variables may change the diagnosis even within the same data set. Thus, as a general observation, analysts should assess the problem for the particular model and the particular data they are using.

Nevertheless, the basic point remains: in contrast to often-expressed concerns about attrition, for many estimates the coefficients on standard variables in equations are unaffected by attrition. This is the case for longitudinal samples for developed countries, and we have shown it to be the case for longitudinal samples in developing countries as well, using a wide variety of outcome variables. Thus, even when attrition is fairly high, as it is in the samples we used, attrition apparently is *not* a general and pervasive problem for obtaining consistent estimates. This suggests that demographers, as well as other social scientists, proceed with greater confidence in their growing attempts to use longitudinal data to control for unobserved fixed factors and to capture dynamic relationships.

6. Acknowledgements

Notes

1. We concentrate here on an approach that has been employed in the econometric literature. Other approaches to the attrition problem are employed in the wider statistical literature. See, for example, Cochran (1977) and Little and Rubin (1987) for further discussions of these alternative approaches.

2. For simplicity in terms of notation and discussion (but with no substantive implications) we assume here that attrition, once it occurs, is permanent. That is, that one respondents drop out of the sample, they do not re-enter. This is the case, of course, for those who drop out of the sample due to mortality and, for the most part, due to permanent migration – and is the case on which the literature focuses. But if there is, for example, circular migration (e.g., see note 9 below on "reverse attrition" in Kenya), individuals may re-enter the sample after dropping out.

3. The analysis of attrition in the above context is therefore slightly different from the issues addressed in the statistical literature on missing values (e.g., Little and Rubin 1987) or non-response (e.g., Ahlo 1990), which is primarily concerned with the case when (a subset of) the dependent or explanatory variables for a respondent are missing at only one or a few survey waves.

4. This is likely, but not guaranteed, because the bias due to observables may be partially offsetting biases due to unobservables, so removing the former may actually increase the biases in the estimates. But, unless there is a reason for a specific presumption that the biases due to the observables is offsetting the biases due to the unobservables, in a probabilistic sense it is likely that lessening the former will lessen the overall attrition bias.

5. The proof relies on the fact that the initial survey sample, which may be a random sample of the population or a sample that is stratified based on time invariant characteristics, changes only through the attrition process. Most panel surveys, including those used in this paper, fall into this category.

6. It is of course possible that the attrition probability is also influenced by time-varying variables that are unobserved at time t due to attrition, and these variables can obviously not be included in the estimation of the weights in Eq. (5). In most applications, however, variables that are observed at time t, such as time-invariant variables, lagged time-variant variables and variables that do not require a completed interview, measure an important subset of the determinants of attrition. Accounting for these factors can therefore substantially reduce attrition biases even if other variables that are unobserved at time t due to attrition also directly affect the attrition probability.

7. The methods applied in this paper do not ordinarily test for or adjust for potentially selective non-response in the initial survey wave. The same restriction applies to all other attrition tests that rely on data collected within the panel survey. Testing for potential biases in the initial survey wave requires that data on the attritors in the initial wave, that is at time $t=1$, are available from other sources such as for instance register information.

8. Fitzgerald et al. (1998) provide a more detailed discussion of the relation between the BGLW test and a direct estimation of the attrition probability based in Eqs. (2 – 3). Because this discussion also provides an intuition for the statistical rationale of the BGLW test, we present it here. In particular, consider a version of the latent attrition index in Eq. (2), where the probability of attrition after the initial survey wave depends linearly on the observed variables x_1 and y_1, where y_1 represents the auxiliary variable z_t in our earlier discussion:

$$A^* = \delta_0 + \delta_1 x_1 + \delta_2 y_1 + v_1 \qquad (7)$$

$$\begin{aligned} A &= 1 \text{ if } A^* \geq 0 \\ &= 0 \text{ if } A^* < 0. \end{aligned} \qquad (8)$$

By inverting Eq. (7), taking expectations and applying Bayes's Rule it can be shown that

$$E(y_1 \mid A, x_1) = \int y_1 f(y_1 \mid x_1) \, w(A, y_1, x_1) \, dy_1 \qquad (9)$$
where

$$w(A, y_1, x_1) = \Pr(A \mid y_1, x_1) / \Pr(A_1 \mid x_1) \qquad (10)$$

which are essentially the inverse of the weights in Eq. (6). The primary difference is that the weights in Eq. (10) are calculated for attritors ($A_t = 1$) and non-attritors ($A_t = 0$). Equation (9) shows that if the weights all equal one, the conditional mean of y_1 is independent of A and hence A will be insignificant in a regression of y_0 on x_1 and A (the conditional mean of y_1 in the absence of attrition bias is $\beta_0 + \beta_1 x_1$, so a regression of y_1 on x_1 will yield estimates of this equation). A noted earlier, the weights in Eq. (10) will equal one only if y_1 is not a determinant of attrition A conditional on x_1. Thus, the BGLW method is an indirect test for the same restriction as the direct method of estimating the attrition function in Eq. (7) itself.

However, if the weights do not equal one, an explicit solution for Eq. (9) in terms of the parameters in Eq. (7) is usually not possible. This solution would require conducting the integration shown in (9). It would be simpler to just estimate a linear

approximation of Eq. (9) by OLS, as is done in the BGLW test. In the linear approximation, the BGLW test therefore determines the magnitude of the effect of *A* on the intercepts and coefficients of the equation for y_1 as a function of x_1. If this effect is significant, it indicates that the conditional mean of y_1 in Eq. (9) depends on *A*, which in turn indicates that the weights in Eq. (10) are not all equal to one and that the variable y_1 is a relevant determinant of attrition.

It should be kept in mind that this BGLW test is not an independent test of attrition bias separate from the test based on the direct estimation of the attrition probabilities in Eq. (7 – 8). It is only a shorthand means of deriving the implications of attrition for the magnitudes of differences in the initial value of the dependent variable y_1 conditional on x_1 between attritors and non-attritors.

9. These households were stratified into three subsamples: (P) (40 percent of the total), which is a stratified random sample of households with children attending PIDIs in which first the PIDI sites were selected randomly and then children within the sites were selected randomly. (A) (40 percent of the total), which is a stratified random sample (based on the 1992 census) of households with children in the age range served by PIDI living in poor urban communities comparable to those in which PIDI had been established, but in which PIDI programs had not been established as of that time. (B) (20 percent of the total), which is a stratified random sample (based on the 1992 census) of households with at least one child in each household in the age range served by PIDI and living in poor urban communities in which PIDI had been established and within a three block radius of a PIDI but without children attending PIDI.

10. There also is "reverse attrition" in the sense of respondents who were present in Kenya 2 but not in Kenya 1: 12 percent (of the Kenya 2 total) for men, 11 percent for women, and 19 percent for couples.

11. These data are not available for 22.4 percent of the men and 21.8 percent of the women interviewed in Kenya 1 but not in Kenya 2.

12. The PSLSD has been alternatively referred to as the SALDRU survey, the South African Integrated Household Survey, and the South African Living Standards Measurement Survey.

13. In practice certain key individuals in the household were pre-designated for tracking if they had moved; in some cases this led to split households in 1998, but that does not affect this analysis which, except for the attrition indicator, uses only 1993 information (May et al. 2000). Figures presented in this paper differ slightly from May et al. (2000) due to updated information on attrition in the sample.

14. There are 1,006 African and Indian children in KwaZulu-Natal in 1993 with complete height, weight, and age information but the following are dropped from the analysis: 23 because the absolute value of at least one of the three height-for-weight z scores, weight-for-age z scores, or weight-for-height z scores exceeded 9.9; 57 who were less than 6 months old; and 29 who were more than 72 months old. If only those re-interviewed as residents (living in the household more than 15 out of the past 30 days) are considered, attrition rises to 30 percent, but the results reported on here are qualitatively the same.

15. This is true at the 10 percent level as well for all of these except for the fine motor ability test score.

References

Alderman, Harold and Jere R. Behrman, 1999, "Attrition in the Bolivian Early Childhood Development Project and Some Tests of the Implications of Attrition," Philadelphia: University of Pennsylvania, mimeo.

Ahlo, Juha M., 1990, "Adjusting for Non-response Bias Using Logistic Regression," *Biometrika* 77(3): 617-624.

Ashenfelter, Orley, Angus Deaton, and Gary Solon, 1986, "Collecting Panel Data in Developing Countries: Does it Make Sense?" LSMS Working Paper 23, Washington, DC: The World Bank.

Becketti, Sean, William Gould, Lee Lillard, and Finis Welch, 1988, "The Panel Study of Income Dynamics after Fourteen Years: An Evaluation," *Journal of Labor Economics* 6: 472-92.

Behrman, Jere R., Kohler, Hans-Peter and Watkins, Susan C. (2001). How Can We Measure the Causal Effects of Social Networks Using Observational Data? Evidence from the Diffusion of Family Planning and AIDS Worries in South Nyanza District, Kenya. Max Planck Institute for Demographic Research, Rostock, Germany, Working Paper #2001-022 (available at http://www.demogr.mpg.de).

Behrman, Jere R., Yingmei Cheng and Petra Todd, 2001, "Evaluating Pre-school Programs when Length of Exposure to the Program Varies: A Nonparametric Approach," Philadelphia: University of Pennsylvania, mimeo.

van den Berg, Gerard J. and Maarten Lindeboom, 1998, "Attrition in Panel Survey Data and the Estimation of Multi-State Labor Market Models," *The Journal of Human Resources* 33(2): 458-478.

Cochran, William G. (1977). *Sampling Techniques*. New York: Wiley & Sons.

Falaris, Evangelos M. and H. Elizabeth Peters, 1998, "Survey Attrition and Schooling Choices," *The Journal of Human Resources* 33 (2): 531-554.

Fitzgerald, John, Peter Gottschalk, and Robert Moffitt, 1998, "An Analysis of Sample Attrition in Panel Data," *The Journal of Human Resources* 33 (2): 251-99.

Foster, Andrew and Mark R. Rosenzweig, 1995, "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," *Journal of Political Economy* 103 (6): 1176-1209.

Grosh, Margaret and Paul Glewwe, eds., 2000, *Designing Household Survey Questionnaires for Developing Countries: Lessons from Ten Years of LSMS Experience*, Oxford, UK: Oxford University Press for the World Bank

Kohler, Hans-Peter 2001, "On the Taxonomy of Attrition in Panel Data: Comments on Fitzgerald, Gottschalk and Moffitt (1998)," Rostock, Germany: Max-Planck Institute for Demographic Research, Mimeo.

Lillard, Lee A. and Constantijn W.A. Panis, 1998, "Panel Attrition from the Panel Study of Income Dynamics," *The Journal of Human Resources* 33 (2): 437-57.

Little, Roderick J. A. and Rubin, Donald B., 1987, *Statistical Analyses with Missing Data*, New York: Wiley.

Maddala, G. S., 1983, *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Maluccio, John A., 2001, "Using Quality of Interview Information to Assess Nonrandom Attrition Bias in Developing Country Panel Data," *Review of Development Economics* (forthcoming).

May, Julian, Michael R. Carter, Lawrence Haddad, and John A. Maluccio, 2000, "KwaZulu-Natal Income Dynamics Study 1993-1998: A Longitudinal Household Database for South African Policy Analysis," *Development Southern Africa* 17(4): p. 567-581.

Powell, J., 1994, "Estimation of Semi-Parametric Models." In R. Engle and D. Mcfadden (eds.), *Handbook of Econometrics, Vol IV*, Amerstdam and New York: North Holland.

PSLSD, 1994, Project for Statistics on Living Standards and Development: South Africans Rich and Poor: Baseline Household Statistics, South African Labour and Development Research Unit, University of Cape Town, South Africa.

Renne, Elisha P., 1997, "Considering Questionnaire Responses: An Analysis of Survey Interactions," Paper presented at the annual meeting of the African Studies Association, Columbus, Ohio, 13-16 November 1997.

Smith, James P. and Duncan Thomas, 1997, "Migration in Retrospect: Remembrances of Things Past," Santa Monica, CA: Rand Labor and Population Program, Working Paper Series 97-06.

Thomas, Duncan, Elizabeth Frankenberg, and James P. Smith, 1999, "Lost But Not Forgotten: Attrition in the Indonesian Family Life Survey," RAND Labor and Population Program Working Paper Series 99-01, Santa Monica, CA: RAND.

Zabel, Jeffrey E., 1998, "An Analysis of Attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an Application to a Model of Labor Market Behavior," *The Journal of Human Resources* 33 (2): 479-506.

Ziliak, James P. and Thomas J. Kniesner, 1998, "The Importance of Sample Attrition in Life Cycle Labor Supply Estimation," *The Journal of Human Resources* 33 (2): 507-3

Appendix

The following is the proof of relation (5) taken from Fitzgerald et al. (1998). Let $f(y_t, z_t|x_t,)$ be the complete-population joint density of y_t and z_t and let $g(y_t, z_t \mid x_t, A_t=0)$ be the conditional joint density. Then

$$g(y_t, z_t \mid x_t, A_t=0) \quad = \quad g(y_t, z_t, A_t=0 \mid x_t) / \Pr(A_t=0|x_t)$$

$$= \quad \Pr(A_t=0|y_t, z_t, x_t)\, f(y_t, z_t|x_t,) / \Pr(A_t=0|x_t)$$

$$= \quad \Pr(A_t=0| z_t, x_t)\, f(y_t, z_t|x_t,) / \Pr(A_t=0|x_t)$$

$$= \quad f(y_t, z_t|x_t,) / w(z_t, x_t)$$

where the third equality follows from the definition of selection on observables in relation (4) and the term $w(z_t, x_t)$ is defined in Eq. (6) in the text. Hence,

$$f(y_t, z_t|x_t,) = w(z_t, x_t)\ g(y_t, z_t \mid x_t, A_t=0).$$

Integrating both sides over z_t gives Eq. (5) in the text.