Southern Africa Labour and Development Research Unit

An evaluation of the determinants and implications of panel attrition in the National Income Dynamics Survey (2008 – 2010)

by

Nic Baigrie and Katherine Eyal

Working Paper Series

Number 103

About the Author(s) and Acknowledgments

Doctoral Student and Lecturer in the School of Economics, University of Cape Town.
Corresponding author: katherine.eyal@uct.ac.za.

The authors would like to thank the National Income Dynamics Survey for funding, Ingrid Woolard for her support and referral in the final stages, Natasha Curry for her invaluable editing and Salma Kagee, Rowan Clarke and Sam Day for helpful comments and input.

Recommended citation

Baigrie, N., Eyal, K. (2013). An evaluation of the determinants and implications of panel attrition in the National Income Dynamics Survey (2008 – 2010). A Southern Africa Labour and Development Research Unit Working Paper Number 103. Cape Town: SALDRU, University of Cape Town

---

ISBN: 978-1-920517-44-1

Working Papers can be downloaded in Adobe Acrobat format from www.saldru.uct.ac.za.
Printed copies of Working Papers are available for R15.00 each plus vat and postage charges.

Orders may be directed to:
The Administrative Officer, SALDRU, University of Cape Town, Private Bag, Rondebosch, 7701,
Tel: (021) 650 5696, Fax: (021) 650 5697, Email: brenda.adams@uct.ac.za

An evaluation of the determinants and implications of panel attrition in the National Income Dynamics Survey (2008 – 2010)[1]

Nic Baigrie and Katherine Eyal[2]

SALDRU Working Paper Number 103
University of Cape Town
July 2013

Abstract

Panel surveys offer a valuable tool for researchers to explore the dynamics underlying individual and household behaviours. The Achilles heel of panel data is attrition. This paper examines the determinants and implications of attrition in the first two waves of South Africa's National Income Dynamics Survey (NIDS). Multivariate tests in labour market and health specifications show that there is some moderate evidence of attrition bias in estimated coefficients based on the non-attriting sample. This bias can be seen in labour market specifications, in particular for men, and for Africans, and to a much lesser degree in health specifications, in particular for small samples of Whites. Researchers should take care when using the panel data set to generalise to the overall population.

Keywords: attrition bias, panel surveys, South Africa, selection on observables, selection on unobservable.
JEL: J10, C23, C24, D00

---

1 SALDRU (2008, 2010).

2 Doctoral Student and Lecturer in the School of Economics, University of Cape Town. Corresponding author: katherine.eyal@uct.ac.za. The authors would like to thank the National Income Dynamics Survey for funding, Ingrid Woolard for her support and referral in the final stages, Natasha Curry for her invaluable editing and Salma Kagee, Rowan Clarke and Sam Day for helpful comments and input.

1. Introduction

The recent proliferation of panel surveys in developing countries offers a valuable tool for social scientists and policy-makers alike to explore the dynamics underlying individual and household behaviours (Nicoletti, Peracchi, & Foliano, 2011). Panel attrition between waves is a potentially severe threat to the value of the data (Thomas et al., 2010), and as such is an on-going concern in any panel survey. This paper examines the sources and consequences of attrition in the first two waves of South Africa's National Income Dynamics Survey (NIDS), the first nationally representative panel survey in South Africa.

This paper examines the detection of selective attrition and attrition bias in the NIDS, with some comparison between panel attrition in developing and developed countries. Exploratory analyses of attrition in simplified health and labour market models are provided, to illustrate the potential implications of attrition bias in these general areas of research. Testing includes bivariate and multivariate examinations of the selective attrition dynamics in the NIDS. The structural equations considered include a canonical Mincerian earnings function, following Magruder & Nattrass' (2006) investigation into attrition in the Khayelitsha Panel Study (KPS), and the adaptation of Wittenberg's (2011) model estimating the correlates of body mass index (BMI)[3] in the NIDS Wave 1.

The trend among the majority of the comparable panel-attrition research from other developing countries, and indeed within some small panels in South Africa, is that while substantial selective attrition is present between waves, the resultant bias in structural estimations is minimal. This is known as the "neutrality of attrition" (Lee, 2003). NIDS is the first nationally representative panel in South Africa however, and whether this result still holds is of interest, as we do not expect the nature of the selection bias to necessarily be the same compared to other countries.

The structural equations we consider show some evidence of attrition bias in estimates for the general population, and for certain sub populations.

Section 2 contains a relevant literature review, section 3 a discussion of the methodology and data, section 4 contains descriptive statistics (bivariate analysis) and discussion thereof, section 5 discusses the results of the multivariate analysis, using the selection on observables method, with some extensions and caveats, and section 6 concludes.

2. Literature Review

2.1 Panel Surveys and Attrition

The increased popularity of panel surveys, both large and small[4], since the mid-1960s (Das, Toepoel, & van Soest, 2011), makes the analysis of the determinants and effects of attrition very necessary, if the many advantages[5] of longitudinal data are to be fully exploited. Panel surveys allow for the

---

[3] Weight (in kilograms)/height (in cm) squared.

[4] Kassirye and Ssewanyana (2010) note that the recent abundance of randomised experiments, examining development effectiveness, in Latin America and Sub-Saharan Africa also provides a wealth of small panel surveys (Behrman & Hoddinott, 2005; Miguel & Kremer, 2012).

[5] Panel surveys often include a wealth of data on socio-economic behaviours and dynamics at both individual and household level (Baird, Hamory, & Miguel, 2008). Common examples include the measurement of demographic transitions, labour market participation and household composition and income sources (Magruder & Nattrass, 2006).

repeated measurement of variables over time, thus addressing the common weaknesses of cross sectional surveys, such as measurement error and omitted variables bias (OVB). Cross sectional surveys fail to capture the unobserved time-invariant heterogeneity which drives a substantial portion of OVB (Maluccio, 2004, Ashenfelter, Deaton and Solon, 1986), and the use of panel data can address these issues. However the threat of attrition to panel samples, defined as a failure to successfully re-interview targeted households or individuals (Kasirye & Ssewanyana, 2010), can reduce the value of panel data. Due to migration, mortality or non-response between waves, attrition can critically undermine the validity of research undertaken using panel data (Magruder & Nattrass, 2006).

The crux of the attrition bias problem, as explained by Falaris (2003), is whether or not respondents who drop out are behaviourally different to those who remain in a panel. The extension of this is whether findings based on a reduced panel can be applied to the entire initial survey population, including those who dropped out.

In cases where large portions of the sample have attrited between waves, this does not necessarily imply the presence of attrition bias. Behrman and Watkins (1998) give an extensive summary of studies which, although significant evidence of selective attrition is found, the authors fail to reject the hypothesis that attrition bias does not affect the slope intercepts in a regression of the structural equation.

For illustration, consider a particular study of the effect of depression on obesity which collects a sample representative of the national population in Wave 1, but in Wave 2 omits to sample those interned in mental institutions or rehabilitation clinics. This systematic loss of a subset of the population causes bias in the study's findings (Dave, Tennant, & Colman, 2011). However, a study aiming to investigate the role of 13th cheque bonus incentives on labour productivity, using the same dataset, could conceivably have no significant bias in its findings. This is because the processes through which the sample selection bias operates are incidental to the analysis and exogenous to the particular structural equation. This illustrates the specificity of attrition bias to the particular dataset and research question involved.

The incentives study in the example above would have incurred attrition which is termed "MAR", where data has gone missing at random. This is often inconsequential for the study's analysis, although a smaller sample size will necessarily reduce the precision or confidence with which statistical relationships can be established (Lee, 2003). This cumulative loss of sample size can be devastating for longer term studies if measures are not implemented to regain new respondents in each wave to compensate for the loss (Lemay, 2009). Doing this without jeopardising representativeness is not easy and Maluccio (2004) notes that the deterioration of "current period representativeness" may be accelerated in societies which undergo more rapid change associated with growth and development, particularly developing countries such as South Africa.

2.2 Determinants of Attrition

Attrition can arise from fieldwork errors, non-contact[6], refusal or death, of which the last three are the primary factors of attrition (Alderman, Behrman, Kohler, Maluccio, & Watkins, 2001; Gray, Campanelli, Deepchand, & Prescott-clarke, 2012). This paper excludes fieldwork errors (which are

[6] Non-contact can be due to failure by the survey to track movers, migration out of the scope of the survey (e.g. emigration, incarceration, conscription), or simply not being at home when the fieldwork team arrives, despite repeated visits.

survey dependent) from the analysis. Much of the discussion which follows is related to how these three factors present in developing economies.

Most surveys which track movers use phone and postal records[7] extensively in their efforts to locate targeted respondents (Lee, 2003). When these networks are significantly less developed, tracking is far more costly[8] (Hill 2004; Falaris, 2003). Some surveys drop sample respondents who have moved between waves from the sample completely, due to the higher costs of tracking migrants, and do not track them at all. These are known as “rooftop” surveys (Lee, 2003), and are discussed below.

Rapid socioeconomic growth and development result in long distance rural-to-urban migration at both the household and individual level (Maluccio, 2004). A higher incidence of upward socio-economic mobility results in respondents who are harder to track, and potentially attrition which operates through substantively different mechanisms to those of more stable economies. However Thomas et al (2001) note that attrition in the Indonesian Family Life survey arises from both ends of the income distribution. Lower income levels make respondents more susceptible to instability in employment and health status (Baird et al., 2008), and increase vulnerability to the social and economic transitions and shocks[9] which are likely to increase attrition propensity (Lee, 2003).

Non-contact is the dominant reason for both individual and household non-response (Lipps, 2009) in almost all developing country surveys, not just those which choose a “rooftop” methodology.

Refusal to respond is determined by a range of factors. There are documented cultural[10] aspects (Maluccio, 2004), as well as the effects of matching between the enumerator and respondent on aspects such as gender, race (Ashenfelter, Deaton, & Solon, 1986), and, more recently, political interests (Pickery, Loosveldt, & Carton, 2001) and religion (Baird et al., 2008). Where high unemployment is present, time may not be a binding constraint (Lee, 2003), and thus refusal rates may be low (Moffit, Fitzgerald, & Gottschalk, 1999).

Developing countries generally have a lower life-expectancy. High mortality rates make it more likely that a respondent will die between waves, *ceteris paribus*. In addition, in countries with flawed death records, migrants are more likely to be erroneously recorded as dead (Ashenfelter et al., 1986). Death and non-contact disproportionately affect developing countries.

In addition to attrition, other factors can hamper the ability to generalise our results – some of these are discussed below.

2.3 Generalisation of Results

Surveys in some countries may not be nationally representative, possibly due to cost constraints (Kasirye and Ssewanyana, 2010). The value of generally applicable policy analysis findings is

[7] The establishment of communications infrastructure makes tracking movers easier, and reduces the probability of non-contact (Maluccio, 2004).

[8] There are many anecdotes of fieldwork teams in the NIDS Wave 2 attempting to relocate respondents whose Wave 1 address is recorded as, for example, “past the third house, turn left and it is the house next to the pond”.

[9] These include job loss, marriage dissolution, family formation and household migration. For young adults, these may also include gaining independence, finishing school and becoming sexually active (Lee, 2003, Dallimore et al., 2002).

[10] Such as the South-East Asian courtesy bias, where respondents refuse to answer in the negative or give answers which might offend the interviewer (Jones, 1963).

particularly high in low income countries, where the implementation of poverty- and inequality-reduction measures is most urgently required (Kasirye & Ssewanyana, 2010; Norris, Richter, & Fleetwood, 2007).

Developing country surveys may often choose a rooftop methodology where movers are not tracked.[11] The cost advantage to these surveys can be considerable.[12] Even among surveys which do track movers, there can be substantial cumulative attrition between waves.[13]

Surveys may be "piggybacked" onto existing surveys which were originally intended to be cross-sectional, one example of which is the Khayelitsha Panel Study (KPS) (Magruder & Nattrass, 2006). This is problematic if the original survey contains insufficient contact information for respondents, is poorly designed or is ill-suited to being a panel survey, creating difficulty in establishing continuity between Waves 1 and 2 (Kasirye & Ssewanyana, 2010).

Loss of sample size reduces the statistical power of the dataset to make precise estimates on parameters of interest. Rooftop surveys at the fieldwork level sometimes maintain sample size, despite loss due to non-response, by "substitution" where missing households are replaced by alternative non-sampled units. This worryingly popular method (Lacerda, Ardington, & Leibbrandt, 2008) implicitly assumes that the non-response mechanism is missing completely at random (MCAR). If the MCAR assumption is violated, the bias in population parameter estimates can be substantial (Rubin & Little, 2002). Some panel studies also include "split-offs", which are newly born individuals or split families who are absorbed into the panel according to selection probabilities, formulated specifically to maintain representativeness[14] (Ashenfelter et al., 1986).

Rindfuss et al. (2004) argue that study designs should take the differential attrition dynamics into account in order to further capture and analyse the processes underlying them, particularly around the issues of mobility and major life-transitions. Falaris (2003) concurs, stating that the observation of attritors is extremely valuable in its own right and that preventing attrition bias is perhaps among the weakest of the reasons to build these dynamics into future studies. Baird et al. (2008) also argue strongly that, in terms of representativeness, the benefits of tracking to data quality heavily outweigh the additional cost. Falaris (2003) and Baird et al (2008) highlight that tracking has previously had remarkable success rates.[15]

In almost all studies, panel attrition is found to be highly selective on individual and community-level socioeconomic characteristics, although this has a far more significant effect on subsequent estimations in most developing country studies (Lee, 2003), where the neutrality of attrition is less

[11] Examples of these include the Indian Additional Rural Incomes Survey (ARIS), the Bolivian Integrated Child Development Program (PIDI), the Kenyan Ideational Change Survey (KDICP), Cameroon's Institut de Formation et de Recherche Demographiques (IFORD) and Vietnam and Peru's World Bank Living Standards Measurement Surveys (LSMS's) (Lee, 2003).

[12] Kasirye and Ssewanyana (2010) estimate that the per-person re-interview cost for those who had not moved was $54, a figure which was more than three times less than the figure for those who had moved, which was $179. This disparity can be attributed to the decision to track even those who had moved to neighbouring countries.

[13] The Michigan Panel Study of Income Dynamics (PSID) had relatively low annual attrition of 8-12%, but has lost over 50% of the original sample over 21 years of surveying.

[14] This practice is common in the Michigan PSID.

[15] In the Indonesian Family Life Survey (FLS) tracking reduces the attrition rate by two-thirds (Duncan Thomas, Frankenberg, & Smith, 2001).

common. Migration and overall attrition rates occasionally result in cases of bias in un-weighted estimations (Kasirye & Ssewanyana, 2010).

2.4 Attrition in Developed Countries

Many aspects differentiate panel attrition in developed and developing countries, though similar attrition rates can be achieved, with some developing country studies achieving remarkably low rates. The Indonesian Family Life Survey, for example, managed to resurvey 94% of their Wave 1 sample in Wave 2. Some developed country surveys experience large sample losses, such as the University of Southern California's Longitudinal Study of Generations which had a 43% attrition rate between two waves (Miller & Wright, 2012). It is hypothesised however that the mechanisms through which attrition operates may differ starkly between the two types of nations.

In more developed countries, higher income levels result in a higher opportunity-cost of time to which Maluccio (2004) attributes the generally higher prevalence of refusal in developed country panel surveys. To the extent that non-response patterns do differ systematically between developing and developed countries, these differences are often associated with outcomes of interest (Lee, 2003).

In developed countries panel attrition occurs more frequently among minorities and younger individuals and households, particularly those at the lower end of the income distribution (Lee, 2003). This suggests an element of economic stress-related migration. Additionally, there is a higher incidence of refusals among those at the upper end of the income distribution, and in comparison to developing countries, endorsing Maluccio's (2004) theory of the opportunity cost of time (Lee, 2003).

2.5 Attrition in South African panel surveys

The 21% individual attrition rate between NIDS Waves 1 and 2 can be compared to the rates of 16% (J. Maluccio, 2000) and 29.5% (Magruder & Nattrass, 2006) for the KwaZulu Natal Income Dynamic Study (KIDS), which took place over ten years in three waves, and Khayelitsha Panel Study (KPS), two waves over four years, respectively (the latter excluding movers).[16] In the KIDS, which is limited to only Africans and Indians, 9% of non-response was due to refusals (Maluccio, 2004). The corresponding figure from the NIDS is 5.4% (authors' own calculations), although Africans are the only race in the NIDS for whom non-contact is more prevalent than refusal. Attrition rates for other races in the NIDS are shown in column 5 of Table A.3 – Part 1. Comparisons of attrition rates between panel surveys are often tricky: the figures above are at the individual level, but the definitions for household level non-response are often inconsistent between surveys (Lee, 2003).

Magruder and Nattrass (2006) provide a labour market orientated analysis of attrition in the KPS, finding socioeconomically upward migration is less evident than other Sub-Saharan African studies predict, and main predictors of attrition as follows: lower household and individual income, living in a shack, recent arrival in the household and being single. Despite evidence of selective attrition, very little evidence of bias inherent in labour market analysis is found.

[16] The annual attrition rate is discussed below, calculated as per Alderman et al (2001). These simple rates are used for the purposes of comparison with other studies.

Maluccio (2004) performs a similar analysis for the KIDS, using selection on un-observables (discussed below). Households in wealthier communities are more likely to have moved between waves in the KIDS, indicating some degree of socially upward migration, which is accompanied by evidence of bias in household expenditure analysis (Maluccio, 2004). Dallimore et al. (2002) find that life-transitions increase the likelihood of attrition among youth in KwaZulu Natal's Transitions to Adulthood study (T2A). They also find a high refusal level among Whites,[17] which results in an attrition rate among Whites in the T2A of 61% (Lee, 2003). The corresponding figure for NIDS is 54% (authors' own calculation). This finding, consistent among South African surveys, poses concern for researchers attempting to study the behaviour of the subpopulation of Whites in South Africa (Lee, 2003).

There appear to be a complex range of effects contributing to attrition in South African surveys. Some are common to most developing countries, but others may be specific to South Africa due to its unique cultural history and the legacy of the Apartheid regime (May, Carter, Haddad, & Maluccio, 1999). This legacy has created immense disparities between races (Branson & Wittenberg, 2007), and the nation has incredibly high levels of both poverty and inequality (Argent, Finn, & Leibbrandt, 2009; Finn, Leibbrandt, & Woolard, 2009).

3. Methodology and Data

The recent literature on panel attrition follows Fitzgerald, Gottschalk and Moffit's (1997) canonical "FGM model", which adapts Heckman's (1979) selection bias model by relaxing the assumption of selection occurring on "un-observables" and allowing attrition to operate through observed variables. The term "selection on observables" has since evolved to include the current definition of an attrition mechanism determined by observed but endogenous variables. Selection on exogenous observables is easily dealt with, as OLS estimators remain unbiased and consistent (Wooldridge, 2009). Under the selection on observed but endogenous variables case, of which lagged outcome variables are a prime candidate, using correction weights based on an estimated attrition function can provide consistent estimators with weighted least squares (WLS), as illustrated below (Vandecasteele & Debels, 2007).

3.1 Model formalisation

The model draws heavily on Fitzgerald, Gottschalk and Moffit (1997). We are interested in the conditional population density $f(y|x)$ of the structural equation, where y is the scalar outcome variable, and x is a vector of independent variables. Working at the population level, and ignoring sampling considerations, A is an attrition dummy, equal to 1 if the individual is missing a Y value due to attrition, and equal to 0 if not. A temporary assumption for simplicity is that X is observed regardless of attrition, which occurs if it is time-invariant, or lagged[18].

[17] This is a common phenomenon in South African panel surveys, which is hypothesised to be due to a combination of the prevalence of secure compounds and gated households, making refusal more common (Lee, 2003), and a higher opportunity cost of time in wealthier households (Maluccio 2004).

[18] Consider for example a standard Mincerian earnings function. Y is Earnings, x is a vector of earnings determinants, particularly educational attainment and years of experience (defined as age – 6 – years of total

The non-attriting panel allows estimation of some function $g(y|x, A = 0)$, but inference of f from g necessitates restrictions (Fitzgerald, Gottschalk, & Moffitt, 1997).

We are interested in the probability function $\Pr(A = 0|x, y, z)$, where z is any observed variable, distinct from x, but ancillary to the structural equation of interest.

In the probability function, y is observed if and only if $A = 0$. If an individual's propensity to attrite is dependent not only on the observed x and z, but also on the partially unobserved y or another unobserved variable (e.g. innate ability), selection is referred to as based on un-observables. When the attrition propensity is independent of the partially unobserved y in Wave 2, selection is deemed to be on "observables" [19] (Moffit et al., 1999). The attrition function: $\Pr(A = 0|x, z)$ represents the probability of attrition conditional only on the observed variables x and z.

Selection on observables occurs when:

(1) $$\Pr(A = 0|x, y, z) = \Pr(A = 0|x, z)$$

If (1) fails to hold, the attrition probability function is dependent on y and therefore cannot be reduced as in equation (1) from the probability function. Maluccio (2004) states the assumptions parametrically in the following canonical one-period selection model, where bold characters represent vectors:

(2) $$y_i = \boldsymbol{x}_i'\boldsymbol{\beta}_1 + e_i \qquad , y_i \text{ observed iff } A_i^* \leq 0$$

(3) $$A_i^* = \boldsymbol{x}_i'\boldsymbol{\beta}_1 + \boldsymbol{z}_i'\boldsymbol{\gamma} + v_i$$

Equation (2) is the structural equation and, y_i is only partially observed, for those for whom the latent index, A_i^*, is negative and who thus remain in the panel. A_i^* is not practically observed, but rather its realisation: whether or not an individual is resurveyed $(A_i = 0)$ (Maluccio, 2004).

The extension of this selection bias model to a two-wave panel attrition application is straightforward and is shown by treating y_i as the Wave 2 outcome of a panel dataset (Maluccio, 2004). We simply use the vector of observed Wave 1 independent variables as the time-invariant x_i and recast equation (3) as the attrition function. The two potential selection mechanisms are then defined as follows:

Selection on un-observables occurs if there is correlation of the error terms of the structural equation (2) and attrition function (3): $Corr(e_i, v_i) \neq 0$ (Maluccio, 2004).

Selection on observables occurs if $Corr(e_i, v_i) = 0$, but $\boldsymbol{z}_i$ is correlated with e_i, and hence is both auxiliary and endogenous to the structural equation (2). Potential $\boldsymbol{z}_i$ include any variable observed in Wave 1 which satisfies criteria, possibly including lagged y_i (Maluccio, 2004).

Estimation of equation (2) ignoring equation (3) yields inconsistent estimates of $\boldsymbol{\beta}_1$ if either selection case applies. This is referred to as "attrition bias" (Maluccio, 2000).

---

schooling, and can be considered to be a lagged variable, as this is a measure of experience from wave 1). For those who attrite between Wave 1 and 2; A = 1.

[19] In the Mincerian example in footnote 5, occupational industry is an often-observed variable, which is excluded from the structural equation (in the sense that a standard earnings function would not be conditioned on industry). However, if a global recession occurs between waves causing many bankers to lose their jobs, and Wave 2 is undertaken by analysing pay-slips, industry is then a determinant of attrition propensity. As such, selection would be on observables, potentially biasing estimations based on the full panel.

The fundamental concern, regardless of which selection case applies, is correlates of e_i. Thus for each particular research question and structural equation, there is a different set of independent variables in equation (2) and e_i thus inevitably changes.

As shown by Moffit, Fitzgerald and Gottschalk (1999), for selection on observables the complete population density can consistently be estimated as follows:

(4) $$f(y,z|x) = g(y,z|x, A = 0)w(z,x)$$

The right hand side is the product of the conditional joint density for the non-attriting sample, and a weighting w. Integration of (4) with respect to z yields the desired outcome: equation 5, the complete population density.

(5) $$f(y|x) = \int_z \; g(y,z|x, A = 0)w(z,x)dz$$

The integration is expanded in Moffit et al. (1999), together with the formulation behind the weight: "w" in equations (4) and (5) is a normalised inverse selection probability, shown in equation (6) below:

(6) $$w(z,x) = \{\Pr(A = 0|z,x)/\Pr(A = 0|x)\}^{-1}$$

3.2 Detection and correction procedures

3.2.1 Selection on un-observables

Fitzgerald et al. (1997) suggest an implicit validation test for attrition bias due to selection on un-observables, by comparing the cross-sectional marginal distributions from an outside data source[20] to those in the panel survey. This is not possible using the NIDS data, as the most likely candidate would be the 2011 Census, which is only scheduled for release in March 2013.

Behrman and Watkins (1998) consistently estimate $\boldsymbol{\beta_1}$ (from equations 2 and 3) in the selection on un-observables case by estimating the attrition index (3). Identifying $\boldsymbol{\beta_1}$ requires an exclusion restriction, that z is validly excluded from the structural equation. Using Heckman's (1979) two-step model to correct for selection, where the first stage estimates the attrition function using all exogenous x_i variables and identifying instruments z_i (Maluccio 2004). This yields a selection correction factor $\widehat{\lambda_i}$ (Maluccio 2004), included in the structural equation, which "controls" for attrition bias. This instrumentation is often hard to rationalise, as attrition-related characteristics would often appear to be included in $\boldsymbol{x}$, with the exception of identifying variables external to the individual (Behrman & Watkins, 1998). In addition, finding valid instruments is often difficult in practice. The parametric method is robust to selection on both observables *and* un-observables (Maluccio, 2004), unlike the weighted least squares (WLS) method discussed below. However more restrictive assumptions[21] (Falaris 2003; Horowitz & Manski 2012) are imposed than in WLS, and

[20] Appropriate sources include any dataset which is representative of the same population to which the panel wishes to generalise its findings, and suffers less attrition, or attrition of a different form.
[21] These assumptions are on the joint distribution of the un-observables in the attrition and structural equations.

other practical difficulties arise[22]. For these reasons, we choose to focus on the selection on observables case in the analysis which follows.[23]

3.2.2 Selection on observables

Selection on observables is corrected using WLS. In this the criteria of the z_i are entirely different: any z_i which affects attrition propensity and is endogenous to Y (related to the density of y conditional on x) would potentially be valid (Behrman & Watkins, 1998). If the structural equation does not include a lagged value of the dependent variable, then y_{t-1} is an ideal example of an applicable z_i in the selection on observables case, assuming it has some predictive power for attrition.

The detection and correction methodology used in the majority of the literature is relatively standardised for the selection on observables case (Kasirye & Ssewanyana, 2010; Lee, 2003). There are two sufficient conditions for the absence of attrition bias on observables: the corrective weights w sum to 1 or the elected z_i are independent of $(y|x)$ (Behrman & Watkins, 1998). Specification tests can therefore be based on either of these conditions.

Moffit et al. (1999) suggest the Hausman test for the significance of the difference in the coefficients of WLS and OLS respectively. To detect whether the attrition mechanism does in fact operate through the chosen observable z, the attrition function (equation 3 above) is estimated using a probit model. In a two-wave model, this involves regressing A on x and z, (in this case y_0) (Fitzgerald et al., 1997). The interpretation of this test is twofold: a significant coefficient on y_0 indicates that it (as z) is indeed a determinant of A, but the implications of this for y are not directly discernible (Moffit et al., 1999).

Becketti, Gould, Lillard and Welch (1988) developed the BGLW test which provides these implications directly: the regression of y_0 on x and A_i (the inverse of the attrition function mentioned in (3), a formal derivation of which is given in Gottschalk et al.'s (1997) paper). The coefficient on x in this equation is neither an unbiased nor consistent estimator of x's coefficient in the attrition function (Moffit et al., 1999), but the implication for our Wave 1 outcome variable is directly apparent as the coefficient on A_i, which is the desired result. This BGLW method is therefore simply a “shorthand” version of the attrition function method, used to directly discern the implications of attrition on the outcome variable (Fitzgerald et al., 1997).

It is important to note that the solution of controlling for the attrition propensity for the un-observables case is very different to that of the observables case, where the weighted least squares method is used with inverse selection probability weights. Simply controlling for attrition as is appropriate for the un-observables case would distort the conditional distribution of y on x. The moment of interest is $E(y|x)$ not $E(y|x,z)$ (Behrman & Watkins, 1998). This illustrates z's role as endogenous, but excluded from the structural equation.

This paper focusses purely on the selection on observables case. The benefits are that, as well as making fewer restrictive assumptions, the method uses the full information where available for the attritors, and not just the fact that they attrited (Falaris, 2003).

[22] Maluccio (2004) also notes that although the Wave 2 outcome variables can be modelled for the non-attriting sample, the model assumptions dictate that they can be modelled using only Wave 1 explanatory variables and fixed effects. This is despite Wave 2 explanatory variables being available for that non-attriting sample.

[23] An extended discussion of the selection on un-observables case is provided in appendix item A.1.

3.3 Introduction to NIDS

NIDS is part of a 2006 initiative by the South African Presidency to "track and understand the shifting face of poverty" (SALDRU & DRA, 2008). The study focusses on household dynamics and general wellbeing, as well as offering a vehicle for the assessment of social-policy efficacy (SALDRU & DRA, 2008). The first wave of fieldwork was conducted in 2008, with Wave 2 being broken down into two phases, in 2010 and 2011. The implementation of the second phase was an explicit attempt to reduce the unit non-response in phase 1, which had been exceptionally high (Brown, Daniels, De Villiers, Leibbrandt, & Woolard, 2012). Phase 2 of Wave 2 focused on re-attempting to interview households which had not been contacted in phase 1, or had refused to be interviewed. The "mop-up" was a success, and the final attrition rate was 21.26% (Brown et al., 2012). Alderman et al.'s (2001) formula for annual attrition rate r is $r = 1 - (1 - q)^{1/t}$, where q is the overall attrition rate, and t the number of years covered by the panel. The resultant annual attrition rate of 11.26% in the NIDS could be considered as mid-range for developing country panel studies, as it is higher than those found in India, Malaysia and Indonesia, but lower than Bolivia, Kenya and Nigeria (Alderman et al., 2001). However this annual rate is much higher than those found in studies in Ethiopia, India, Peru and Vietnam, which had annual rates (over 7 years) of 0.7%, 0.5%, 0.6% and 0.3% respectively (Outes-Leon and Dercon, 2008, Young Lives, 2012). The KPS and KIDS surveys mentioned above have annual rates of 1.72%, and 8.37% respectively. NIDS is comparatively high compared to these two South African surveys.

An overview of attrition in the NIDS [24] is provided in the beta release user manual (Brown et al., 2012), which we consolidate into a transition matrix for wave 1 and 2 respondents, (in appendix item A.2). From wave 1 to wave 2, there were 2136 refusals (approximately 37% of attritors), 2714 respondents who were not contacted (48%), and 846 deaths (15%) (Brown et al., 2012). It is important to note that the nature of attrition may change, and attrition may increase substantially the more waves are added to the panel. In particular, the size of certain sub-populations in the panel such as Indians/Asians may be decimated, resulting in the inability to generalise results for these groups. Due to data constraints this paper only examines the issues relating to short term attrition between 2 waves. In the period between waves there were many events which could affect patterns of mobility, as well as income dynamics, not least of all the "global recession" which sparked worldwide concern. Whether these factors systematically affected attrition in the NIDS, or other contemporary panel surveys, is a potential avenue for further research.

3.4 Tracking and Identification

Tracking is often accompanied by other attrition-reduction measures intended to reduce the incidence of non-response. These include the provision of pecuniary incentives for respondents – up to $60 in the National Longitudinal Survey of Youth (NLSY) and $80 in the Michigan PSID (Laurie & Lynn, 2008). Conversion strategies are also quite common, where "conversion specialists" aim to sway the decision of respondents who have refused, either telephonically or in an entirely personalised letter, by addressing their concerns and emphasising their importance to the survey. This conversion method reduced refusals by up to 50% in some waves of the NLSY (Lee, 2003), and is also used in the NIDS.

[24] Defined as a failure to re-interview wave 1 respondents in wave 2.

NIDS follows individuals, and not households[25]. Household members are defined as those who have lived in the household for at least 15 days in the past year, have shared common food, as well as resources, and contributed to shared resources (Brown et al 2012). Resident household members usually reside at the house for at least 4 nights in the past week. Residents are defined as either continuing sample members (CSMs[26]) or temporary sample members (TSMs[27]).

Only CSMs from wave 1 are tracked between waves, and surveyed in each wave. TSMs, (individuals living with CSMs) in waves 2, 3 etc. are also surveyed, but not tracked. NIDS follows individual CSMs rather than household heads. SALDRU makes use of a panel maintenance system, and various incentives for the field work company, and for panel respondents, to minimise unit non-response, and ensure individuals were successfully tracked after moving. These include multiple visits per household where individuals are not at home on first visit, non-conditional gifts to respondents, and where necessary, additional phases of data capture (phase 2 in wave 2) to decrease the attrition rate arising in main from moving and refusals. Where non-response occurs, these cases are referred back to field head-quarters, and other members of the original household are contacted, in an effort to locate movers. GPS location details are also used in a bid to verify contact details.

Where households split between waves, every effort is made to track and survey both households, if both contain CSMs. New household identifiers are allocated where households either split, or if the entire household moves, as household identifiers are associated with geographical locations. If CSMs cannot be located in a particular wave, attempts are made in further waves to make contact again.

The bivariate analysis presented below introduces the detection of selective attrition and attrition bias in the NIDS, and exploratory analyses of attrition bias in simplified health and labour market models.

4. Descriptive Statistics

4.1 Bivariate analysis

4.1.1 Reasons for Attrition

Attrition between waves 1 and 2 of the NIDS is examined in Appendix item A.3. Africans are the only race for whom non-contact is more common than refusal, making up 54% and 28% of their total attrition respectively. Africans have by far the lowest attrition rate, of 18.05%, but account for over two-thirds of the total attrition in the population, as they comprise 80% of the total NIDS sample.

---

[25] Thus individuals are identified by a unique person identifier (pid), and household identifiers are not identical across waves (Brown et al, 2012).

[26] All Wave 1 resident household members, children included, and any children of female CSMs in the following waves (biological or adopted) (Brown et al, 2012).

[27] TSMs are not CSMs, but reside with a CSM at the time of the interview (Brown et al, 2012).

Indians and Asians exhibit the highest ratio of refusal to non-contact, with the former nearly four times more likely. Whites have a slightly lower ratio than Asians and Indians, and also the lowest mortality rate. Whites have the highest attrition rate, with more than half of those targeted failing to be resurveyed. There is no substantial differential attrition behaviour across gender. This is possibly because the majority of attrition occurs at the household level, dwarfing any possible gender-driven effects at the individual level. Wittenberg (2011) finds that (specifically in health-related aspects) there are substantial differences between genders within each race, and we follow his method of providing health specifications for males and females separately in table 7, for Africans and Whites.

The bivariate distribution of attrition with age displays significant differences across age categories, which suggest differential mechanisms through which selective attrition operates (Cullen & Levitt, 2006). Mortality predictably increases consistently with age[28] after the expected initial spike in deaths among infants (Preston, 1980). The largest share of attrition is among 15-35 year olds, which is consistent with the literature. The attrition rate among this age group is slightly lower than that of over 65 year olds, 44% of whom died between waves. It is only for those above 35 that refusal outweighs non-contact.

There is a higher attrition rate amongst urban respondents, for whom refusal is more common than non-contact (see Table A.3, part 2). For rural respondents, non-contact is more common than refusal. In rural areas, the communication infrastructure with which fieldworkers attempt to relocate respondents may be less developed; however, the proportion of respondents not contacted is in fact lower for the rural subpopulation.

Panel E of A.3 part 2 shows the distribution of attrition by income[29] decile. The income findings lend support to the correlation of refusal propensity with income. Refusal is more common than non-contact only among the wealthiest 20% of respondents. This is driven by both a declining probability of non-contact, and increasing incidence of refusal with income (see columns 1 & 2). These dynamics combine to yield a refusal-to-non-contact ratio that ranges from 0.4 in the lowest income decile, to almost 5 times higher for the top income decile. There is also a substantial drop in known between-wave-mortality among the upper tiers of the income distribution.

Appendix A.4 shows the reasons for attrition broken down by province. The Western Cape and Gauteng have the highest per-capita attrition rates of 28% and 25% respectively, although they account for only a quarter of total attrition. The Eastern Cape and KwaZulu Natal are the only provinces for which non-contact dominates refusals. They also account for over 40% of the total sample and their effect therefore drives the overall attrition pattern: over the total population, non-contact is 30% more prevalent than refusal. KwaZulu Natal, despite having the second lowest attrition rate (13%), accounts for the largest share of total attrition.

[28] The AIDS-related increase in deaths among young adults is not apparent with such wide age-brackets (Wittenberg & Collinson, 2005), as the majority of AIDS-related deaths presumably occur *within* the 15-35 bracket, and not *between* age brackets.

[29] The income variable used is the household's per-adult income, with an adult equivalence weighting of infants counting for one third of an adult, and children a half. Overall, mean household size is 5.5 (extrapolated from Appendix A.6), or 4.5 using the adult equivalence scale (Atkinson, Rainwater, & Smeeding, 1995), which is more likely to reflect the reality that income is disproportionately allocated between members.

4.2 Bivariate detection discussion

Further analysis into the specific traits and differences among those lost to refusal, non-contact or death, is helpful to detect any systematic attrition which could drive bias in later research – this is provided in Table 1.

4.2.1 Attrition

"Refusers"[30] are significantly older, more educated, wealthier and come from bigger, better-off and older households (see Table 1). Those lost to non-contact are less likely to be self-employed, married, obese, or suffer from hypertension.

Attritors are older, more educated, more likely to be male, richer, and come from smaller households than non-attritors. More attritors are, on average, employed as wage-earners, and fewer are unemployed. They are also more likely to have been married, and of those who have married, attritors are more likely to have divorced than non-attritors. They have, on average, lower BMI[31], although more are classified as overweight[32]. Interestingly, it is also evident that attritors suffer more limitations in Activities of Daily Living (a higher ADL score), and higher depression levels. However from column 3 we see that these findings are driven primarily by their prevalence among attritors who are known to have died between waves.

These findings are all significant at the 95% level, although this evidence of systematic attrition does not necessarily imply attrition bias in any estimation. The magnitude of some of the differences is important, particularly those of education and household, individual and labour income. Those who refuse have on average between one and two years more education than those who don't attrite, and also more than the overall average (column 6). Those who refuse earn more than double the average earnings of the rest of the population, and come from households which are similarly twice as wealthy as their non-refusing counterparts, and thus likely to have more educated members.

When replicating table 1 for Africans[33], similar patterns are found. The only additional difference is a greater obesity rate among non-attritors.

---

[30] Pair one of the t-tests compares the means of each variable between those who refused (column 1), and those who were not successfully contacted (column 2).
[31] We make use of the BMI, obesity and hypertension methods given in Ardington and Case (2009).
[32] The apparent contradiction is due to the mutually exclusive categories of "overweight" and "obese". There is no difference in obesity, and slightly fewer overweight non-attritors. Attritors have a lower average BMI.
[33] Available from the authors on request.

Table 1: Attritor Characteristics by Type of Attrition						
Unconditional Means	Refused	Non-Contact	Deceased	Attritors	Non-Attritors	Total
	(1)	(2)	(3)	(4)	(5)	(6)
Age	**30.1**	25.3	47.7	**30.4**	26.1	27.0
Years of Education	**7.85**	6.14	5.07	**6.62**	5.64	5.85
% Male	0.48	0.49	0.47	**0.49**	0.44	0.45
HH Income	**9808**	4164	3467	**6177**	4104	4544
HH Size	**4.82**	4.65	5.30	4.81	**5.74**	5.55
HH No. of Children	1.67	**1.92**	1.96	1.83	**2.48**	2.34
HH No. of Adults	**3.14**	2.72	3.33	2.97	**3.26**	3.20
Individual Income	**2369**	1512	939	**1748**	937	1109
Labour Income (a)	**1810**	1011	293	**1175**	722	828
% Wage Employed	0.33	0.36	0.13	**0.31**	0.23	0.24
% Casual Employed	0.04	**0.06**	0.04	0.05	0.05	0.05
% Self Employed	**0.09**	0.04	0.05	0.06	0.05	0.06
% Unemployed	0.67	0.64	0.87	0.69	**0.77**	0.76
% Married	**0.37**	0.22	0.27	0.28	0.28	0.28
% Widowed	0.06	0.05	0.22	0.09	0.09	0.09
% Divorced	0.04	0.03	0.03	**0.04**	0.02	0.03
% Never Married	0.45	**0.56**	0.38	0.48	**0.52**	0.51
BMI	25.8	25.3	24.5	25.4	**26.0**	25.8
Perceived Health Index	3.77	3.74	2.74	3.57	3.54	3.55
% Overweight	0.23	0.20	0.20	**0.21**	0.19	0.20
% Obese	**0.21**	0.17	0.17	0.18	0.19	0.19
ADL Score	0.36	0.35	1.61	**0.54**	0.43	0.46
Depression Index	11.3	11.1	14.8	**11.7**	10.4	10.7
% Hypertension	**0.36**	0.28	0.48	0.35	0.35	0.35
N	2136	2714	846	5696	21098	26794

Notes: Labour Income is for those employed only.
Table 1 presents the means of these Wave 1 variables according to type of attrition for all attritors, non-attritors, and the total Wave 1 sample. Two pairs of T-tests are also presented for differences in the means. The first is a t-test between "refusers" and "non-contacters". The second is between all attritors and all non-attritors. In both cases, for each variable, if the means of the two sub-populations are significantly different at the 95% level, then the greater of the two means is **bold**. The t-test assumes normality of the distributions of each variable, which is patently inappropriate for dummy variables such as gender and marital status, although the results serve nevertheless as indicators of whether there is a substantial difference between the respondents in question. This analysis serves merely as a cursory overview of the selection mechanism, and as such, biases in the t-tests are not of major concern for our later analysis.

4.2.1 Mobility

The Wave 2 data includes a variable which indicated whether or not a respondent had moved between waves, or stayed in the same household. Systematic mobility could potentially lead to attrition bias if movers are not tracked. Appendix A.5 presents the means for movers and stayers who attrited, broken down by reason for attriting.

Movers that refuse are younger than stayers, and come from larger households which contain more children. They are less likely to be married, and more likely never to have been married. Younger unmarried individuals are more able to migrate, having fewer commitments and obligations. Movers are generally healthier in terms of obesity, hypertension and ADL score, and these differences are significant. Similar patterns are seen for movers that are not contacted, and these individuals are also from larger, wealthier households than stayers. These patterns are also seen for non-attritor movers versus stayers, and all movers vs. stayers (these results presented in A.6).

Among attritors, movers come from larger households and report lower labour incomes. Among non-attritors, this pattern is reversed. Because almost 80% of the population is present in both waves, it is these non-attritors whose effects dominate the overall population parameters.

Mobility is strongly age-related, and on the whole movers earn more, and come from wealthier, larger households. The evidence of upward mobility is substantial with regard to labour and household income[34], although not so for individual income.[35]

5. Results

5.1 Multivariate analysis and detection findings

If attrition in the NIDS is due to selection on observables, z is any observed variable which is a determinant of attrition propensity, endogenous to $(y|x)$, but is excluded from the structural equation. In the NIDS; y_2 is partially unobserved, although we maintain the simplifying assumption that the $\boldsymbol{x_i}$ and z_i are time-invariant

The results of these detection methods will vary for every specific structural equation (Watkins & Warriner 2003) . The structural equation (2) is estimated using health and labour market specifications.

The attrition function – equation (3) - regresses A_2, whether a respondent attrited, on $\boldsymbol{x_i}$ and z. The z in our case is a lagged value of Y, so all right-hand side variables are measured in Wave 1 (Fitzgerald et al., 1997). This is analogous to the formulation used by Brown et al. (2012) in constructing the attrition-corrected weights for Wave 2, they however estimate the probability of Wave 2 attrition conditional on Wave 1 characteristics, but these characteristics are restricted to only age, race, educational attainment and marital status[36]. Where appropriate, post-stratified sampling weights are used to obtain generally representative estimates.

---

[34] Individual income includes pensions, grants, unemployment payments, inheritances and other miscellaneous income sources.

[35] Naturally, this correlation doesn't imply causality. The trends do however support an upward mobility theory.

[36] How many control variables to include is debatable: a parsimonious estimation may well be desirable, but a more comprehensive estimation may provide weights which better capture the full dynamics of attrition.

The attrition function is estimated for the total population and for Africans only (in tables 2 and 3), due to the differential attrition patterns for Africans. The sample excludes those under 15 years of age, for whom the attrition reason (e.g. a decision to move, or household refusal) was likely exogenous, made on their behalf and thus beyond their control. Following Moffit et al. (1999), we choose to exclude the known dead from the attrition probits, and all further analysis – in order to ensure comparability of the results to other papers.

We follow the methodology of Moffit et al. (1999), including as the Wave 1 labour market outcome variables: labour income and its quadratic, and a dummy variable indicating the lack of any labour income. Moffit et al. (1999) estimate an expanding probit specification initially on the outcome variables alone. Standard Mincerian determinants of labour income are included as $\boldsymbol{x_i}$, and finally other controls, including household level socio-economic indicators, geographic variables, and an indicator of the likelihood of moving in the near future. We use similar specifications.

Kasirye and Ssewanyana (2010) perform the probit estimation with dual outcome variables, household expenditure and the shock of rebel attacks in Uganda, by including both outcome variables in the probit. We combine their method with Moffit et al.'s (1999) model to formulate the three expanding specifications, initially regressing attrition solely on z (the outcome variables y_0), then controlling for their standard determinants. The final regression includes other household- and individual-level socioeconomic controls[37].

Table 2 presents an estimation of the attrition function for the total population[38]. Model 1A and 1B (not reported) are, following Moffit et al.'s (1999) model specifications, probits on each of the outcome variable sets individually. Model 2 controls only for the two outcome sets together, and hypertensive status is the only variable which is not extremely significant (all others are significant at most at the 2% level). Hypertension does however increase in significance in the expanded models, while labour income squared is no longer significant once further controls are introduced in models 3 and 4 (similarly for Africans in table 3).

Model 3 includes standard determinants for the outcome variables, including the classical Mincerian regressors (Mincer, 1974) used by Magruder and Nattrass (2006), as well as the contributing factors and correlates of health status as used by Wittenberg (2011)[39].

The quadratic relationship between attrition and labour income, found by Moffit et al. (1999), is evident in model 2 of Table 2, where attrition probability is low in the middle of the income distribution (marginal effects reported are evaluated at the variable's mean), and higher at extrema. This significance is attenuated however in models 3 and 4, once further controls are added. This implies perhaps that once other factors are taken into account, such as race, education, health etc., (all determinants of labour income), this quadratic effect is split up into its constituent parts, and thus is no longer significant by itself. This is also seen in table 3 for Africans. Those who do not earn income from labour are however substantially less likely to attrite, an effect which is significant in all three models.

[37] NIDS collects a broad array of socioeconomic information. We include in $\boldsymbol{x_i}$ only the standard controls, for comparability with other developing country attrition studies which have fewer variables at their disposal.

[38] When these estimates are run using only household heads in Moffit et al (1999), to remove the possibility of intra-household correlation affecting the results, the neutrality of attrition mainly holds for earnings functions, except the intercepts of such functions. We re-run the specifications in table 6 and 7 clustering on hhid, and find no changes in either the coefficients or p values, which is re-assuring. These results are available from the authors on request.

[39] In this case, the expanding specification are not testing for any "optimal" model, but rather depicting how existing coefficients change once certain controls are added, F-tests for joint significance are thus omitted.

Of the controls introduced in model 3, we see that Coloureds, Indian/Asians and Whites are substantially more likely to attrite than Africans (the base case). This provides further justification of separate analysis of Africans in table 3. The significance of the coefficient on Coloureds, wanes once Model 4's controls are introduced.

Model 4 illustrates the attrition function, with the attrition dummy regressed on both outcome variables, as well as a full set of controls for their standard determinants[40] and individual and household level socio-economic indicators. Living in a rural area lowers attrition probability by 5%, and the quality of the respondent's prime water source raises it very slightly. The largest significant coefficient in model 4 is on White. Whites have an attrition probability that is 23% higher than Blacks – in line with the descriptive statistics in table 2.

Table 2: Attrition Probit on Total Population in the NIDS

Dependent Variable = Attritor	Model 2		Model 3		Model 4	
	δP/δx	P>|z|	δP/δx	P>|z|	δP/δx	P>|z|
Labour Income(a)	0.00	**0.00**	0.00	0.54	0.00	0.33
No Labour Income	-0.09	**0.00**	-0.10	**0.00**	-0.07	**0.00**
Labour Income^2	0.00	**0.02**	0.00	0.25	0.00	0.20
BMI status	-0.02	**0.00**	-0.02	0.37	-0.01	0.25
Hypertension	-0.01	0.43	0.02	0.13	0.02	0.24
Male			0.02	0.09	0.02	0.28
Coloured			0.09	**0.01**	0.03	0.34
Indian/Asian			0.13	**0.02**	0.17	**0.00**
White			0.26	**0.00**	0.23	**0.00**
Age			0.05	0.30	0.03	0.61
Age^2			0.00	0.52	0.00	0.93
Years of Education			-0.06	0.28	-0.03	0.59
Smoker			0.04	**0.02**	0.03	0.09
Married					-0.02	0.32
HH per adult Income					0.00	0.55
HH Size					0.00	0.30
Rural					-0.05	**0.02**
Water Quality Index					0.02	**0.05**
Reading – Home Lang.					0.00	0.96
Constant	0.42	0.00	1.65	0.12	1.09	0.37
Pseudo R^2	0.02		0.05		0.08	
N	12,773		12,699		12,271	

Notes: Labour Income measured in '000's of Rands. Sample restricted to respondents older than 15 years, and excludes known dead. P-values are **Bold** where coefficient is significant at the 95% significance level. A full set of Provincial dummies is included in specification 4. In all

[40] Model 4 also includes unreported controls, which include province dummies, and number of children. Other insignificant controls are mentioned in the table notes.

three specifications, those not earning a labour income are less likely to attrite. Indian/Asians and Whites are more likely to attrite than Africans. Those living in a rural area are less likely to attrite. Many variables were not significant in model 2, 3 and 4, and certain of these have been excluded from the display. These include ADL Score, Depression Index, Perceived Health, HH number of children, Toilet Quality Index, Refuse Removal, Reading – English, Writing – Home Language and Writing – English.

When the attrition mechanism among Africans is investigated separately in table 3, we see that the differential attrition among those not earning labour incomes remains significant, decreasing attrition probability by around 7%. BMI's significance attenuates in model 3 and 4 (although the coefficient was never large to begin with), while smoking, which was not significant in the overall models for the full sample in table 2, is associated with an increase in the probability of attrition of 5%. This is an interesting finding; Wittenberg (2011) found smoking to be a strong (inverse) correlate of BMI, which in turn was shown, among the African population, to be a legitimate marker for well-being.

For Africans, measures of age, education, well-being, household size and literacy are not significant. Quality of water source access has a small positive significant coefficient.

The model explains very little of the overall variation in attrition among Africans and among the total population. This is illustrated by the Pseudo R-Squared statistics reported in the final row.

Moffit et al. (1999) include in their regressors a variable which indicates an individual's likelihood of moving in the future, there is however no comparable variable available in NIDS. Fitzgerald, Gottschalk and Moffit (1997) find substantial effects of marital status on attrition propensity, which are emulated here, with marriage being significant for both the overall population, and Africans alone. Table 4 repeats the estimation of model 4, the most comprehensive estimation of the attrition function, by income tercile among Africans. It is clear that different attrition dynamics are present at the different income levels, although many variables are not significantly different across the terciles (specific variables mentioned in the table notes). In the poorest tercile, those living in the Eastern and Northern Cape and the Free State are more likely to attrite, and these effects are large (between 13 and 20%). Poor Africans living in Mpumalanga are 11% less likely to attrite. Province has no impact on the richest individuals. Those living in the Western and Eastern Cape in tercile 2 are more likely to attrite. Among poorer Africans, higher toilet quality reduces the probability of attrition by 2%, as does having more children in the household, although a larger household over all increases the probability of attrition, but this effect is small. For middle and upper terciles, increased household size reduces the probability of attrition.

Table 3: Attrition Probit on Africans in the NIDS						
Dependent Variable = Attritor	Model 2		Model 3		Model 4	
	δP/δx	P>|z|	δP/δx	P>|z|	δP/δx	P>|z|
Labour Income(a)	0.00	0.06	0.00	0.07	0.00	0.51
No Labour Income	-0.07	**0.00**	-0.07	**0.00**	-0.05	**0.01**
Labour Income^2	0.00	0.07	0.00	0.10	0.00	0.20
BMI status	-0.02	**0.00**	-0.01	0.26	-0.01	0.14
Hypertension	-0.03	0.05	0.01	0.65	0.00	0.96
Male			0.01	0.35	0.00	0.84
Age			0.03	0.45	0.00	0.96
Age^2			0.00	0.53	0.00	0.61
Years of Education			-0.03	0.43	0.00	0.94
Smoker			0.05	**0.01**	0.04	**0.05**
Married					-0.02	0.37
Co-habiting					0.06	**0.01**
HH per adult Income					0.00	**0.02**
HH Size					0.00	0.45
Rural					-0.03	0.20
Water Quality Index					0.02	**0.04**
Toilet Quality Index					-0.01	0.07
Reading – Home Lang.					-0.01	0.58
Constant	-0.46	**0.00**	-1.29	0.18	-0.58	0.59
Pseudo R2	0.02		0.03		0.06	
N	10,272		10,213		9,833	

Notes: (a) Labour Income measured in '000's of Rands. Sample restricted to respondents older than 15 years, and excludes known dead. Full set of Provincial dummies included in specification 4.

P-values are **Bold** where coefficient is significant at the 95% significance level.

In all three specifications, those not earning a labour income are less likely to attrite. The significance of BMI as a predictor of attrition wanes once marital status and household-level controls are introduced (Model 4). The variables with the largest marginal effects on attrition propensity are a lack of labour income (negative effect), smoking and cohabiting with a partner (positive effects).

Many variables were not significant in model 2, 3 and 4, and certain of these have been excluded from the display. These include ADL Score, Depression Index, Perceived Health, Divorced, HH number of children, Refuse Removal, Reading – English, Writing – Home Language and Writing – English.

Table 4: Attrition Probits by Income Tercile for Africans in the NIDS						
Dependent Variable = Attritor	Tercile 1		Tercile 2		Tercile 3	
	δP/δx	P>|z|	δP/δx	P>|z|	δP/δx	P>|z|
Labour Income(a)	0.00	0.26	0.00	0.38	0.00	0.32
No Labour Income	0.04	0.48	-0.08	0.12	-0.03	0.37
Labour Income^2	0.00	0.66	0.00	0.46	0.00	0.20
BMI status	0.00	0.79	-0.01	0.24	-0.02	0.27
Male	0.00	0.92	0.03	0.21	-0.03	0.33
Smoker	0.02	0.52	0.05	0.08	0.03	0.36
Cohabiting	0.04	0.23	0.14	**0.00**	0.02	0.70
HH per adult Income	0.00	0.06	0.00	0.21	0.00	0.33
HH Size	0.02	**0.00**	-0.02	**0.00**	-0.02	**0.01**
HH No. of Children	-0.03	**0.00**	0.02	**0.03**	-0.03	0.10
Water Quality Index	0.02	0.08	0.00	0.85	0.06	**0.04**
Toilet Quality Index	-0.02	**0.04**	0.00	0.76	-0.01	0.33
Western Cape	0.03	0.77	0.19	**0.01**	-0.04	0.68
Eastern Cape	0.20	**0.00**	0.11	**0.00**	0.07	0.21
Northern Cape	0.14	**0.03**	0.01	0.87	-0.01	0.93
Free State	0.13	**0.01**	0.07	0.08	-0.06	0.30
Mpumalanga	-0.11	**0.00**	-0.03	0.38	-0.03	0.63
Constant (b)	3.65	0.12	-2.54	0.31	0.48	0.78
Pseudo R^2	0.08		0.10		0.08	
N	3571		3341		2921	

Notes: (a) Labour Income in '000's of Rands. **Bold** indicates significance at the 95% level. (b) Constant reports its probit coefficient, not marginal effect. Unreported controls include full dummy set, and literacy indices. Post-stratified weights used, but do not adjust for attrition. Income terciles are by household per adult income, terciles 1 is poorest, 3 is wealthiest. It is clear that attrition dynamics differ with income amongst Africans in the NIDS. In the poorest tercile, those living in the Eastern and Northern Cape, and Free State are more likely to attrite, and the size of these coefficients is large. Province has no impact on the richest individuals. Those living in the Western and Eastern Cape in tercile 2 are more likely to attrite.

Many variables were not significant, and certain of these have been excluded from the display. These include Hypertension, Age and Age^2, Years of Education, ADL Score, Depression Index, Perceived Health, Married, Divorced, HH number of adults, Refuse Removal and literacy indices.

The BGLW, or inversion test, which Becketti et al. (1988) develop to provide the direct implication on the outcome variables of any selective attrition, is presented in table 5. The test is performed for log of monthly wage and BMI.[41] The test necessarily reiterates the findings of the probit estimates, as it is their inverse (Becketti, Gould, Lillard, & Welch, 1988).

Table 5: "BGLW" Test for Mincerian and Health Specifications Using the Total NIDS Wave 1 Population

Dependent Variable = Ln(Monthly Wage)	(1) Mincerian	
	Coeff.	P>|t|
Female	-0.41	**0.00**
Coloured	0.26	**0.00**
Asian/Indian	0.92	**0.00**
White	0.85	**0.00**
Age	0.11	**0.00**
Age^2	0.00	**0.00**
Years of Education	0.15	**0.00**
Attritor Indicator	0.07	0.09
Constant	0.08	0.36
N	5053	
R^2	0.49	

Dependent Variable = BMI	(2) Health	
	Coeff.	P>|t|
Female	3.40	**0.00**
Coloured	-0.70	0.14
Asian/Indian	0.01	1.00
White	0.35	0.62
Employed	0.81	**0.02**
Age	0.39	**0.00**
Age^2	0.00	**0.00**
Years of Education	0.08	0.08
L(Wage)	0.57	**0.01**
L(HH p/a Income)	0.15	0.52
HH No. Adults	0.06	0.52
HH No. Children	0.04	0.70
Smoker	-2.02	**0.00**
Hypertension	0.81	**0.00**
Depression Index	-0.03	0.31
ADL Score	0.02	0.89
Exercise Index	-0.26	**0.00**
Attritor Indicator	-0.25	0.48
Constant	8.35	**0.00**
N	3636	
R^2	0.26	

Notes: The BGLW, or "inversion" test, includes in the Wave 1 structural equation a variable indicating whether the respondent attrited in Wave 2. The significance of the attrition indicator in either structural equation is the test for attrition bias in the specification concerned. In neither equation is the attrition indicator significant. Attrition bias does not therefore seem to be a major concern in either health or Mincerian specifications on the full NIDS panel.

There is no evidence of attrition bias in either structural equation for the overall population. Other authors report the same findings in their BGLW tests; that despite there being substantial selective attrition in the panel, the resulting bias is negligible (Alderman et al., 2001; Falaris, 2003; Fitzgerald et al., 1997; Magruder & Nattrass, 2006; Moffit et al., 1999). Even substantial selective attrition often does not hamper the generalisability of conclusion based on panels of stayers (Falaris, 2003) particularly in labour market and demographic models.

[41] The similar test for Africans is presented after in the BMI estimates in table 8. It reveals the presence of some bias, which is however not large in size.

5.2 Exploratory extension into structural equations

The extensive selective attrition, often on critical variables, thus accounts only for a small portion of the overall attrition between waves, and therefore yields only minimal bias. In tables 6 and 7, we investigate whether findings using the full panel are in fact representative of both stayers and attritors, comparing the total sample successfully surveyed in Wave 1, to the non-attriting sample in wave 2 (Moffit et al., 1999). The presence of bias will imply doubt as to the generalizability of results using the non-attriting panel.

Basic estimations of structural equations (equation (3)) are run, which include a set of attrition-interaction terms, and the attritor dummy itself. By including these interactions, which are simply the product of the attrition indicator and each variable, it is possible to see how the slope coefficients on each variable would change if the estimation were based on the full panel, for whom the attrition index is equal to zero, or on the total sample, which consists of the panel sample, and also the attritors, for whom the attrition index equals one. Any significant interaction coefficients in these regressions indicate a differential effect between non-attritors and the total population.

In table 6, a basic Mincerian earnings function is estimated (Mincer, 1974), ignoring the potential endogeneity of education. For the purpose of identifying attrition bias, this method is sufficient (Magruder & Nattrass, 2006), although an alternative is to estimate reduced form equations for any potentially endogenous variable (Falaris, 2003); Maluccio, 2004).

Table 6: Attrition Bias in a Mincerian Earnings Function						
Dependent Variable = Ln (Monthly Wage)	Men and Women		Men		Women	
	(1)		(2)		(3)	
	Coeff.	P>|t|	Coeff.	P>|t|	Coeff.	P>|t|
Male	0.45	**0.00**	-	-	-	-
Coloured	0.18	**0.01**	0.11	0.25	0.37	**0.01**
Indian/Asian	0.94	**0.00**	0.83	**0.00**	0.78	**0.01**
White	0.85	**0.00**	0.87	**0.00**	0.65	**0.00**
Age	0.11	**0.00**	0.13	**0.00**	2.53	**0.00**
Age^2	0.00	**0.00**	0.00	**0.00**	0.00	**0.00**
Education	0.16	0.10	0.16	**0.00**	0.12	**0.00**
A*Male	-0.06	0.46	-	-	-	-
A*Coloured	-0.01	0.94	-0.02	0.88	0.00	0.33
A*Indian/Asian	-0.25	0.34	-0.21	0.46	0.04	0.53
A*White	0.04	0.73	0.06	0.74	0.00	0.95
A*Age	0.01	0.70	0.01	0.80	0.04	0.81
A*Age^2	0.00	0.42	0.00	0.44	0.00	0.77
A*Education	-0.03	**0.00**	-0.04	0.01	0.00	0.52
Attritor	0.47	0.25	0.60	0.23	-0.19	0.78
Constant	2.98	**0.00**	3.40	**0.00**	3.16	**0.00**
N	5053		2775		2278	
R-Squared	0.47		0.45		0.50	
Wald/Chi Sq F	135.2		72.2		116	
Joint F	2.14		2.56		0.34	
Joint F: P>|F|	**0.04**		**0.02**		0.92	

Notes: Estimates weighted using calibrated weights, and full dummy set for province included. Wald, or Chi^2 F is for overall model.
"Joint F" statistics are for the full set of Attrition-Interaction terms (A*Variable). **Bold** indicates significance at the 95% level. The attritor dummy is not significant in any of the regressions, but the Joint F test for men and women, and then for men in particular, is significant – there appears to be some attrition bias present here.

Only one interaction term is significant, Attrition*Education, for men and women – the rest are insignificant which seems promising. The attrition indicator is also insignificant in each column. Columns 2 and 3[42] inspect the genders separately. We see the expected coefficients on the variables included for race, age, education as significant predictors of log monthly wage.

[42] It is important to remember not to compare these models directly, with the R-squared for example, as the differing sample sizes (due to different patterns of missing data) mean they cannot be compared directly. This is true for the health specification as well.

Joint significance of the attrition interactions would imply that the slopes of the earnings function differ between attritors and the non-attriting panel. Despite that none of the interaction terms are significant except for one, and that the attrition dummy is not significant, the F tests do show in table 6, for the pooled sample and for men only, that we reject the null hypothesis that the coefficients are jointly equal to zero. This suggests that labour market findings based on the restricted sample in both waves may be biased due to attrition.

Table 7 below provides specifications of a basic health model, based on Wittenberg's (2011) analysis of the correlates of BMI in South Africa. In the pooled sample, the attrition interaction terms are jointly significant. The p-values of the individual variables show that this result is driven largely by gender and race (Asian/Indian specifically). In columns 2 – 5, these are controlled for (Coloureds and Asian/Indians are omitted to allow comparison with Wittenberg's (2011) original table). The attrition-interaction terms are neither jointly nor individually significant for either gender among Africans (as seen in the t and F tests), but are for Whites. This significance could be due to the remarkably low sample sizes in these two regressions, less than one tenth of their African counterparts. The significant variables for White females are years of education, being a smoker or having hypertension, and the interaction of education and attrition. For white men, the natural logs of both wage and household per-adult income, ADL score, and the interaction of the first two are significant. The conflicting signs on the income coefficients however lend weight to the argument that this significance is spurious due to the small sample size, as these variables would hypothetically affect BMI through similar mechanisms. It appears that the joint significance of the interaction terms in the pooled sample could be driven by the significance seen in the sample of Whites, which is relatively small.

Our earlier conclusion of systematic attrition in the NIDS panel, evidenced by significant differences in the means of critical variables between attritors and non-attritors, would seem to conflict with some of the multivariate conclusions above. This is a common occurrence. Cichello, Fields and Leibbrandt (2001) note that much of the existing literature reaches the same conclusion: that "despite these mean differences in attritors' characteristics, structural relationships are not affected" (Cited in Magruder & Nattrass, (2006)). This "neutrality of attrition" (Lee, 2003) is entirely model specific, and there may be biases in other structural equations in the NIDS, particularly studies of mobility dynamics, which are a major correlate of attrition. However, Clarke and Eyal (2012) do not find evidence of attrition bias in these areas.

Table 7: Attrition bias in Correlates of BMI in the NIDS										
Dependent Variable: BMI	Pooled		African-Female		African-Male		White-Female		White-Male	
	(1)		(2)		(3)		(4)		(5)	
	Beta	Sig.	Beta	Sig.	Beta	Sig.	Beta	Sig.	Beta	Sig.
Female	4.12	*	-		-		-		-	
Employed	0.52		0.02		0.23		4.51		-0.42	
Age	0.40	*	0.86	*	0.11		0.62		0.56	
Age^2	0.00	*	-0.01	*	0.00		-0.01		0.00	
Years of Education	0.10		0.34	*	0.01		1.36	*	-0.50	
L(Wage)	0.47	*	0.13		0.90	*	-1.85		-3.60	*
L(HH p/a Income)	0.34		0.27		0.10		-3.08		8.06	*
Smoker	-2.10	*	-1.94		-1.76	*	-5.39	*	-3.35	
Hypertension	0.63	*	0.48		0.40		4.51	*	-0.24	
ADL Score	0.17		0.23		-0.37	*	0.17		-6.95	*
Exercise Index	-0.19	*	-0.37		-0.01		-0.74		-0.35	
A*Female	-2.25	*	-		-		-		-	
A*Asian/Indian	3.75	*	-		-		-		-	
A*Employed	0.55		0.42		0.64		6.19		4.40	
A*Age	-0.07		-0.41		0.23		-0.86		-0.62	
A* Age^2	0.00		0.00		0.00		0.01		0.01	
A*Education	-0.06		-0.43	*	0.09		-2.04	*	0.36	
A*L(Wage)	0.59		1.55		-0.34		4.58		4.82	*
A*L(HH p/a Inc)	-0.72		-0.39		-0.19		-1.72		-7.04	*
A*ADL Score	-0.30		0.56		-0.14		2.79		6.63	
Attritor	0.97		-0.77		-4.11		6.00		19.9	
Constant	7.82	*	1.96		13.0	*	48.6	*	-17.8	
Observations	3636		1164		1391		113		133	
R-Squared	0.27		0.19		0.24		0.65		0.56	
Joint F	1.78		0.87		1.01		2.19		2.18	
Joint F: P>|f|	**0.02**		0.58		0.44		**0.02**		**0.02**	

Notes: Estimates weighted using calibrated weights, and full dummy set for province included.
"Joint F" statistics are for the full set of Attrition-Interaction terms. **Bold** indicates significance at the 95% level and * for coefficients. Significance of interaction terms illustrates potential attrition bias in health related research which uses the full panel without accounting for these attrition dynamics.
The F tests show that attrition bias is potentially a concern in the full sample, and in the two last sub populations.
Many of the variables were not significant in the regressions, and certain of these have been excluded from the display. These include Coloured, Indian/Asian and White, HH number of children, HH number of Adults, Depression Index (all levels), Age and Age^2, Depression Index.

5.3 Caveats and recommendations

While the estimations in this paper are robust to heteroskedasticity and weighted where appropriate to account for clustering in the survey design and yield nationally representative estimates, a potential thorn in the side of investigations into panel attrition is structural endogeneity. Falaris (2003) and Maluccio (2004) include in their labour market exploration reduced form estimations to deal with the endogeneity of education in earnings functions. This is feasible; although Angrist and Pischke (2008) claim that in some situations ordinary least squares (OLS) may be better suited to the estimation of sample weights than the conventional (probits, logits or tobits) limited dependent variable (LDV) estimators. The attrition function, with attrition as a binary dependent variable, has traditionally been estimated with a probit-model, (Brown et al., 2012).

Despite the additional assumptions needed to use non-linear estimators when dealing with endogeneity (Angrist & Pischke, 2008), they are more appropriate for the attrition case than OLS, as fitted probability values are often not near the mean, and marginal effects are therefore more accurately estimated with non-linear estimators. A comparison between the two methods would be better suited to a survey with more than three waves, to include attrition hazard-rate dynamics (Durrant & Goldstein, 2008; Lugtig, 2011).

Wave 2 of NIDS is unfortunately ill-suited to a parametric approach to selection on un-observables, as it lacks suitable instrumental variables or a comparable dataset. The 2011 Census will be released near the release of wave 3 of NIDS, and this would be an exemplary dataset with which to compare the marginal distributions of critical variables (Macurdy, Mroz, & Gritz, 2012). The release of currently withheld information on interviewer and fieldwork details could also provide a valuable opportunity (Hill & Willis, 2012) to estimate a two-step model with identifying instruments (Puhani, 2000). These are two ways in which the analysis provided here could be complemented by examination of selection on un-observables, a more comprehensive methodology (Maluccio, 2004).

Attritors are so different from non-attritors, and we know that race , and gender, matter a great deal when it comes to the neutrality of attrition. Table 6 and 7 show evidence of some concern for attrition bias in both health and labour market specifications, from joint significance tests, despite attrition indicators and interaction terms being mostly individually insignificant. Bias is strongly present for Whites but small sample sizes may make us suspicious of these results. It is possible that our models are either not very well specified (as evidenced by fairly low R-squared values in tables 6 and 7), or a large amount of noise is present in the data. The latter could possibly be due to the fact that this was the first wave of the first nationally representative panel undertaken in South Africa. Comparison between waves 2 and 3 (when available), and 1 and 3, should further our understanding on this point, if data capture issues have been successfully reduced.[43]

In tables 6 and 7, many of the coefficients are fairly different, even if not significantly so. We are interested in whether or not this would make a large impact on the predicted values for attritors or non-attritors. For the Mincerian function, calculations for men and women in column 1, using mean values for the variables, yield no difference in predicted wage[44], between attritors and non-attritors.

A concern with attrition analysis is that non-response at the baseline is not considered. This can result in extreme values for panel weights, and concerns regarding standard attrition analysis methods. The wave 1 weights used where appropriate are intended to account for the survey design and initial non-response, and the wave 2 panel weights account for both survey design and attrition. A problem is when those groups less likely to respond in wave 1 are disproportionately represented among attritors (Brown et al., 2012), such as Whites in their twenties. This may result in very high panel weights – the product of high wave 1 weights, and high attrition weights in wave 2. For this reason panel weights are trimmed in the NIDS (to the 1st and 99th percentile) to ensure accurate estimates for these sub-populations. In NIDS (see A.2), half of wave 1 non responders are found in

[43] Thanks to a referee for this useful comment.
[44] Calculations available from the authors on request, similar results obtained for the health specifications.

wave 2, and it may be useful to compare them to non-attritors in wave 2, or respondents in wave 1, to see quite how different they are. Power could be an issue here, as the sample of those found in wave 2 is fairly small. If they are very different, this could endanger the neutrality of attrition result.

6. Conclusion

The benefits of panel surveys, contentious in the mid 1960's due to uncertainty about their superiority over cross-sectional datasets (Ashenfelter et al., 1986), are now well documented. Attrition poses a threat to the validity and generalisability of findings based on multiple waves of panel data, due to potential bias in structural equation coefficients due to selective attrition. The current literature follows Moffit et al.'s (1999) model of attrition, which posits that the attrition mechanism operates through either observable, or un-observable variables. The method for dealing with the latter follows a two-step estimation (Heckman, 1979), and requires either a comparable dataset, or identifying instruments. Neither of these is available in the case of the NIDS. Selection on observables is however detected, with differential attrition dynamics evident for different races. This selective attrition is generally among younger, healthier and wealthier individuals, in all races.

Multivariate tests for the implications of the selective attrition in the NIDS find that, specifically in research based on labour market or health specifications, there is some moderate evidence of attrition bias in estimated coefficients based on the non-attriting sample (joint significance of the set of interaction terms, but actual attrition indicators and interaction terms not individually significant). However the attrition bias in health seems mainly to be driven by bias amongst Whites only, a very small proportion of the total sample.

Investigations restricted to the subpopulation of Africans find that attrition dynamics differ with income level. Neutrality applies in the structural equations for health, as seen in the BGLW test in table 8 (which agrees with the conclusions from table 7), but not for the labour market specification. The attrition indicator for Africans is significant in the Mincerian regression – attritors have monthly wages which are 11% higher than non-attritors. Researchers using the panel dataset should investigate thoroughly when generalising to the overall population with labour market specifications.

The neutrality of attrition is a common finding, particularly among developing countries (Lee, 2003). Evidence of substantial attrition bias seems to be the exception to the rule, despite extensive selective attrition in many panels (Cichello, Fields, & Leibbrandt, 2001; Lee, 2003). Nevertheless, the existence and implications of attrition bias are very much dependent on the specific structural equation in any research model, and researchers should therefore approach attrition bias on a case by case analysis (Maluccio, 2004).

Table 8: "BGLW" Test for Mincerian and Health Specifications Using Africans Only

Dependent Variable = Ln(Monthly Wage)	(1) Mincerian:	
	Coeff.	P>|t|
Female	-0.45	**0.00**
Age	0.11	**0.00**
Age^2	0.00	**0.00**
Years of Education	0.14	**0.00**
Attritor	0.11	**0.02**
Constant	3.66	**0.00**
N	3327	
R^2	0.36	

Dependent Variable = BMI	(2) Health:	
	Coeff.	P>|t|
Female	4.21	0.00
Employed	0.50	0.20
Age	0.39	**0.00**
Age^2	0.00	**0.00**
Years of Education	0.13	0.01
L(Wage)	0.58	0.02
L(HH p/a Income)	0.16	0.53
HH No. Children	-0.06	0.61
HH No. Adults	0.07	0.52
Smoker	-1.46	0.00
Hypertension	0.60	0.01
Depression Index	-0.02	0.56
ADL Score	0.08	0.67
Exercise Index	-0.12	0.23
Attritor	-0.58	0.14
Constant	8.90	0.00
N	2555	
R^2	0.29	0.00

Notes: In the BGLW test, a Wave 2 attrition indicator is included in the Wave 1 structural equation. Significance thereof is the test for attrition bias in the specification concerned. In the health equation, the attrition indicator is not significant, while in the Mincerian equation it is significant. Attrition bias may therefore seem to be a concern in Mincerian specifications for Africans in the full NIDS panel. Unreported controls in the health regression include province dummies.

References

Alderman, H., Behrman, J. R., Kohler, H.-peter, Maluccio, J. A., & Watkins, S. C. (2001). Attrition in Longitudinal Household Survey Data. *Demographic Research*.

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly Harmless Econometrics : An Empiricist ' s Companion*. Princeton University Press.

Argent, J., Finn, A., & Leibbrandt, M. (2009). Poverty : Analysis of the NIDS Wave 1 Dataset Discussion Paper no . 13.

Ashenfelter, O., Deaton, A., & Solon, G. (1986). *Collecting Panel Data in Developing Countries: Does It Make Sense ?* Washington D.C.

Atkinson, A. B., Rainwater, L., & Smeeding, T. M. (1995). Income Distribution in OECD countries, OECD Social Policy Studies. No. 18. Paris.

Baird, S., Hamory, J., & Miguel, E. (2008). *Tracking, Attrition and Data Quality in the Kenyan Life Panel Survey Round 1 (KLPS-1)*. *Quality* (Vol. 1).

Becketti, S., Gould, W., Lillard, L., & Welch, F. (1988). The Panel Study of Income Dynamics after Fourteen Years : An Evaluation. *Journal of Labour Economics*, *6*(4), 472-492.

Behrman, J. R., & Hoddinott, J. (2005). Programme Evaluation with Unobserved Heterogeneity and Selective Implementation : The Mexican PROGRESA Impact on Child Nutrition. *Oxford Bulletin of Economics and Statistics*, *67*(4).

Behrman, J. R., & Watkins, S. C. (1998). Attrition and Some Tests of the Implications of Attrition. University of Pennsylvania.

Branson, N., & Wittenberg, M. (2007). The Measurement of Employment Status in South Africa using Cohort Analysis, 1994-2004. *South African Journal of Economics*, *75*(June), 1994-2004.

Brown, M., Daniels, R. C., De Villiers, L., Leibbrandt, M., & Woolard, I. (2012). National Income Dynamics Study Wave 2 Beta Release User Manual. Cape Town.

Clarke, R., Eyal, K. (2012). Free To Move: The Microeconomic Determinants of Migration in Post-Apartheid South Africa. Unpublished Manuscript.

Cichello, P. L., Fields, G. S., & Leibbrandt, M. (2001). Are African workers getting ahead in the new South Africa ? Evidence from KwaZulu - Natal , 1993 – 1998. *Social Dynamics: A journal of African studies*, *27*(1), 120-139.

Cullen, J., & Levitt, S. (2006). The Effect of School Choice on Participants: Evidence from Randomized Lotteries - Appendix C: Testing for Attrition Bias. *Econometrica*, *74*(5), 1191-1230.

Das, M., Toepoel, V., & van Soest, a. (2011). Nonparametric Tests of Panel Conditioning and Attrition Bias in Panel Surveys. *Sociological Methods & Research*, *40*(1), 32-56. doi:10.1177/0049124110390765

Dave, D. M., Tennant, J., & Colman, G. J. (2011). Isolating the Effect of Major Depression on Obesity.

Durrant, G. B., & Goldstein, H. (2008). *Working Paper M10 / 08 Analysing The Probability Of Attrition In A Longitudinal Survey* (No. M10/08).

Falaris, E. M. (2003). The effect of survey attrition in longitudinal surveys : evidence from Peru , Cote d ' Ivoire and Vietnam. *Journal of Development Economics*, *70*, 133-157.

Finn, A., Leibbrandt, M., & Woolard, I. (2009). Income & Expenditure Inequality : Analysis of the NIDS Wave 1 Dataset Discussion Paper no . 5.

Fitzgerald, J., Gottschalk, P., & Moffitt, R. (1997). An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics. *Journal of Human Resources*.

Gray, R., Campanelli, P., Deepchand, K., & Prescott-clarke, P. (2012). Exploring survey non-response : the effect of attrition on a follow-up of the 1984-85 health and life style survey. *The Statistician*, *45*(2).

Heckman, J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, *47*(1), 53-161.

Hill, D. H., & Willis, R. J. (2012). Reducing Panel Attrition A Search for Effective Policy Instruments. *Journal of Human Resources*, *36*(3), 416-438.

Hill, Z. (2004). Reducing attrition in panel studies in developing countries. *Access*, *33*(3), 493-498. doi:10.1093/ije/dyh060

Horowitz, J. L., & Manski, C. F. (2012). Nonparametric Analysis of Randomized Experiments With Missing Covariate and Outcome Data. *Journal of the American Statistical Association*, *95*(449).

Jones, E. (1963). The Courtesy Bias in South-East Asian Surveys. *The International Social Science Journal*, *15*(1), 70-76.

Kasirye, E., & Ssewanyana, S. (2010). Impacts and determinants of panel survey attrition: The case of Northern Uganda Survey 2004-2008.

Lacerda, M., Ardington, C., & Leibbrandt, M. (2008). Sequential Regression Multiple Imputation for Incomplete Multivariate Data using Markov Chain. *Human Development*. Cape Town.

Laurie, H., & Lynn, P. (2008). The Use of Respondent Incentives on Longitudinal Surveys. *Statistics*.

Lee, U. (2003). *Panel Attrition in Survey Data: A Literature Review* (No. 41). University of Cape Town.

Lemay, M. (2009). *Understanding the Mechanism of Panel Attrition*. *Doctor*.

Lipps, O. (2009). *Attrition of Households and Individuals in Panel Surveys*. *Social Sciences*.

Lugtig, P. (2011). Panel Attrition . Separating stayers , sleepers and lurkers. *Joop Journal Of Object Oriented Programming*, 1-19.

Macurdy, T., Mroz, T., & Gritz, R. M. (2012). An Evaluation of the National Longitudinal Survey on Youth, *33*(2).

Magruder, J., & Nattrass, N. (2006). Exploring Attrition Bias: The Case of the Khayelitsha Panel Study (2000-2004). *South African Journal of Economics*, *74*(December), 769-781.

Maluccio. (2004). Using Quality of Interview Information to Assess Nonrandom Attrition Bias in Developing-Country Panel Data. *Review of Development Economics*, *8*(1), 91-109.

Maluccio, J. (2000). *Attrition in the KwaZulu Natal Income Dynamics Survey*.

May, J., Carter, M. R., Haddad, L., & Maluccio, J. (1999). *KwaZulu-Natal income dynamics study (KIDS) 1993- 1998 : A longitudinal household data set for South African policy analysis* (pp. 1993- 1998).

Miguel, E., & Kremer, M. (2012). Worms : Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica*, *72*(1), 159-217.

Miller, R. B., & Wright, D. W. (2012). Detecting and Correcting Attrition Bias in Longitudinal Family Research. *Family Relations*, *57*(4), 921-929.

Mincer, J. (1974). Schooling, Experience and Earnings. *National Bureau of Economic Research*, 41-63.

Moffit, R., Fitzgerald, J., & Gottschalk, P. (1999). Sample Attrition in Panel Data : The Role of Selection on Observables. *les Annales d'Economie et de Statistique*.

Nicoletti, C., Peracchi, F., & Foliano, F. (2011). Estimating Income Poverty in the Presence of Missing Data and Measurement Error. *Journal of Business and Economic Statistics*, *29*(1), 61-72. doi:10.1198/jbes.2010.07185

Norris, A., Richter, L., & Fleetwood, S. (2007). Panel Studies in Developing Countries : Case Analysis of Sample Attrition over the past 16 Years the Birth to Twenty Cohorts in Johannesburg, South Africa. *Journal of International Development*, (May), 1143-1150. doi:10.1002/jid

Outes-Leon, Ingo and Stefan Dercon (2008): "Survey Attrition and Attrition Bias in Young Lives". Young Lives Technical Note 5, Available at: http://www.younglives.org.uk/files/technical-notes/survey-attrition-and-attrition-bias-in-young-lives

Pickery, J., Loosveldt, G., & Carton, A. (2001). The Effects of Interviewer and Respondent Characteristics on Response Behavior in Panel Surveys: A Multilevel Approach. *Sociological Methods & Research*, *29*(4), 509-523.

Preston, S. H. (1980). *Causes and Consequences of Mortality Declines in Less Developed Countries during the Twentieth Century* (pp. 289-360).

Puhani, P. (2000). The Heckman Correction for Sample Selection and Its Critique. *Journal of Economic Surveys*, *14*(1), 53-68.

Rubin, D., & Little, R. (2002). *Statistical Inference with Missing Data*. Wiley.

SALDRU, & DRA. (2008). *NIDS Fieldwork Manual*. Cape Town.

Southern Africa Labour and Development Research Unit. National Income Dynamics Study 2008, Wave 1 [dataset]. Version 4.1. Cape Town: Southern Africa Labour and Development Research Unit [producer], 2012. Cape Town: DataFirst [distributor], 2012

Southern Africa Labour and Development Research Unit. National Income Dynamics Study 2010-2011, Wave 2 [dataset]. Version 1. Cape Town: Southern Africa Labour and Development Research Unit [producer], 2012. Cape Town: DataFirst [distributor], 2012

Thomas, D, Witoelar, F., Frankenberg, E., Sikoki, B., Strauss, J., Cecep, S., & Suriastinin, W. (2010). Cutting the Costs of Attrition: Results from the Indonesian Family Life Survey. *Human Development*.

Thomas, Duncan, Frankenberg, E., & Smith, J. (2001). Lost but not Forgotten: Attrition and Follow-up in the Indonesia Family Life Survey. *Journal of Human Resources3*, *Summer*(36-3), 556-592.

Vandecasteele, L., & Debels, A. (2007). Attrition in Panel Data : The Effectiveness of Weighting. *European Sociological Review*, *23*(1). doi:10.1093/esr/jcl021

Watkins, Susan C, & Warriner, I. (2003). How do we know we need to control for selectivity ? *Demographic Research*. doi:10.4054/DemRes.2003.S1.4

Wittenberg, M., & Collinson, M. (2005). Restructuring of households in rural South Africa : Reflections on average household size in the Agincourt sub-district 1992-2003 ∗.

Wittenberg, M (2011). Estimating expenditure impacts without expenditure data using asset proxies, *Economics Letters*, Volume 110, Issue 2, Pages 122-125.Wooldridge, J. M. (2009). *Introductory Econometrics: A Modern Approach* (4th ed.). Cengage Learning.

Young Lives (2012): Young Lives Third Round Country Reports for Ethiopia India, Peru and Vietnam. Available at: http://www.younglives.org.uk/our-publications/country-reports

Appendices:

A.1

Fitzgerald et al. (1997) suggest an implicit "validation" test for attrition bias due to selection on un-observables, by comparing the cross-sectional marginal distributions from an outside data source to those in the panel survey. Poor timing prevents the use of the South African Census, the next wave of which is only scheduled for release in March 2013.

Behrman and Watkins (1998) consistently estimate β1 (from equations 2 and 3) in the selection on un-observables case by estimating the attrition index (3). The identification of β1, however, requires an exclusion restriction. This is that the z mentioned in the model above, which is a determinant of attrition probability, is validly excluded from the structural equation. The selection corrected model is thus estimated according to Heckman's (1979) two-step model, where the first stage is an estimation of the attrition function using all exogenous xi variables and all identifying instruments zi (Maluccio 2004). This yields a selection correction factor λi, which is included in the estimation of the structural equation, and which "controls" for the attrition bias in the model. This instrumentation is often hard to rationalise, as attrition-related characteristics would often appear to be included in x, with the exception of identifying variables external to the individual (Behrman & Watkins, 1998).

Zabel (1998) and Ziliak and Kniesner (1998) follow this parametric method, as does Maluccio (2004), who uses "quality of interview" variables as identifying instruments in the case of the KIDS which could be excluded from most structural equations, and are also partial determinants of attrition.

NIDS includes interview-specific variables, but only regarding duration and details on languages used. Due to the process of interviewer/respondent language and race matching by the fieldwork company, these variables – which could feasibly have served as identifying instruments – do not contain a sufficiently exogenous component of variation (Moffit et al., 1999). Measures of fieldwork quality would be valuable for the application of the parametric selectivity method in NIDS and information on interviewer specific variables could also shed light on the consequences of interviewer-respondent interactions of the sort identified in Deaton and Solon (1986) and Baird et al. (2008). It is worth noting that Behrman and Watkins (1998) point out that the only appropriate identifiers are those external to the individual respondent, and not predetermined in any respondent-interviewer matching process (Vandecasteele & Debels, 2007) in the fieldwork stage.

The above parametric method, using either a suitable identifying instrument or comparable dataset, has a distinct advantage relative to the WLS methodology discussed below in that it is robust to selection on both observables *and* un-observables (Maluccio, 2004). It does, however, impose more restrictive assumptions on the joint distribution of the un-observables in the attrition and structural equations (Falaris 2003; Horowitz & Manski 2012). In addition, finding valid instruments is often hard to do in practice. Maluccio (2004) also notes that although the Wave 2 outcome variables can be modelled for the non-attriting sample, the model assumptions dictate that they can be modelled using only Wave 1 explanatory variables and fixed effects. This is despite Wave 2 explanatory variables being available for that non-attriting sample.

A.2:

Table 1: Wave 1 to Wave 2 Individual Outcomes in the NIDS							
Wave 1	Wave 2						
	Success	Refused / N-Avail.	HH NR	Emigrated	Dead	Total	W1 CSM's
Success	21098	520	4279	51	846	26794	28247
Ref/N-Avail.	952	91	365	2	43	1453	
Not in W1	6591	209	23	0	140	6963	
Total	28641	820	4667	53	1029	35210	

Notes: Of the 28247 continuing sample members (CSM's), for whom contact details were on file after wave 1, 1453 had refused, or were otherwise unavailable in Wave 1. 952 of these were successfully resurveyed in Wave 2. The attritors consist of 520 Refusals, 4279 Household non-respondents, and 51 emigrants who were targeted in Wave 2, but not successfully resurveyed.

A.3

Table A.3 – Part 1: Reasons for Attrition in the NIDS, broken down by Race, Gender and Age.

Sub-population	Refusal (1)		Non-Contact (2)	Deceased (3)	Total Attritors	Sub-Population Attrition Rate	Share of Total Attrition	Total Sub-Population size
Panel A: Attrition Reasons by Race								
African	27.51%	<	53.95%	18.53%	3831	18.05%	67.26%	21223
Coloured	47.67%	>	42.40%	9.93%	1007	26.10%	17.68%	3858
Indian/Asian	74.67%	>	20.00%	5.33%	150	39.47%	2.63%	380
White	69.21%	>	26.84%	3.95%	708	53.11%	12.43%	1333
Panel B: Attrition Reasons by Gender								
Male	37.73%	<	46.94%	15.33%	2929	19.99%	51.42%	14655
Female	37.26%	<	48.39%	14.35%	2767	22.79%	48.58%	12139
Panel C: Attrition Reasons by Age Group								
Infant (0 – 3)	30.29%	<	62.36%	7.35%	449	17.42%	7.97%	2578
Child (3 – 12)	40.35%	<	56.27%	3.38%	798	14.45%	14.16%	5521
Teen (12 – 15)	42.65%	<	53.05%	4.30%	279	14.87%	4.95%	1876
Adult (15 - 35)	36.31%	<	54.30%	9.38%	2046	25.37%	36.32%	8066
Adult (35 - 59)	41.05%	>	39.49%	19.46%	1408	22.49%	24.99%	6260
Pensioner (60+)	33.79%	>	22.17%	44.04%	654	26.92%	11.61%	2429
Total	2136	<	2714	846	5696	21.26%	100.00%	26794

Notes: Columns 1 – 3 report the number and percentage of each sub-population that attrited, by each reason. Panels A, B and C compare Races, Genders and Ages.

Table A.3 – Part 2: Reasons for Attrition in the NIDS, broken down by Geo-Type and Income

Sub-population	Refusal (1)		Non-Contact (2)	Deceased (3)	Total Attritors	Sub-Population Attrition Rate	Share of Total Attrition	Total Sub-Population size
Panel D: By Geo-type								
Urban	46.58%	>	42.19%	11.23%	3330	26.87%	58.46%	12392
Rural	24.73%	<	55.33%	19.95%	2366	16.43%	41.54%	14402
Panel E: By Household (per Adult-Equivalent) Income Decile								
1 - Poorest	24.29%	<	60.53%	15.18%	494	18.40%	8.67%	2685
2	27.35%	<	57.26%	15.38%	468	17.48%	8.22%	2677
3	21.17%	<	59.69%	19.13%	392	14.58%	6.88%	2689
4	32.91%	<	49.04%	18.05%	471	17.65%	8.27%	2669
5	23.95%	<	52.79%	23.26%	430	16.04%	7.55%	2680
6	32.74%	<	47.73%	19.53%	507	18.94%	8.90%	2677
7	29.11%	<	52.26%	18.63%	553	20.64%	9.71%	2679
8	34.63%	<	44.17%	21.20%	566	21.11%	9.94%	2681
9	45.93%	>	44.97%	9.10%	725	27.07%	12.73%	2678
10 - Richest	63.39%	>	31.93%	4.68%	1090	40.69%	19.14%	2679
Total	2136	<	2714	846	5696	21.26%	100.00%	26794

Notes: Columns 1 – 3 report the number and percentage of each sub-population attriting for each reason. Indicators between columns 1 and 2 show that Non-Contact dominates Refusal for income deciles 1 to 8, and those in rural areas. For the top 2 deciles; Refusal is twice as common as Non-Contact, with a 30% - 40% attrition rate.

A.4

Appendix A.4: Reasons for Attrition in the NIDS, broken down by Province

Province	Refusal (1)		Non-Contact (2)	Deceased (3)	Total Attritors	Provincial Attrition Rate	Share of Total Attrition	Provincial Sample Size
Western Cape	67.50%	>	24.43%	8.07%	880	28.36%	15.45%	3103
Eastern Cape	26.91%	<	53.30%	19.79%	773	22.58%	13.57%	3424
Northern Cape	50.46%	>	34.04%	15.50%	329	18.98%	5.78%	1733
Free State	54.18%	>	21.09%	24.73%	275	19.00%	4.83%	1447
KZN	29.00%	<	45.30%	25.70%	969	13.20%	17.01%	7340
North West	48.26%	>	30.03%	21.72%	373	16.70%	6.55%	2233
Gauteng	54.12%	>	35.46%	10.42%	595	25.27%	10.45%	2355
Mpumalanga	48.74%	>	27.31%	23.95%	238	14.04%	4.18%	1695
Limpopo	41.02%	>	37.89%	21.09%	256	10.55%	4.49%	2427
Outside RSA	0.00%	<	100.00%	0.00%	51	63.75%	0.90%	80
Missing	1.57%	<	98.43%	0.00%	957	100.00%	16.80%	957
Total	2136	<	2714	846	5696	21.26%	100.00%	26794

Notes: Columns 1 – 3 report the number and percentage of respondents in each province attriting for each reason. Indicators between columns 1 and 2 show that Refusal generally dominates Non-Contact except in the Eastern Cape and KwaZulu Natal. These provinces have the largest sample sizes, given in the final column, which is why their effects govern the overall attrition pattern: total Non-Contacts (2714) outweigh total Refusals (2136). The second- and third-last rows give the figures for emigrants and those for whom we have no locational information. Western Cape and Gauteng have the highest attrition rates, illustrated in the 5th column, although they account for only a quarter of the total attrition, shown in the second-last column. Eastern Cape and KZN account for nearly a third of total attrition, while the Northern Cape, Free State, North West, Mpumalanga and Limpopo each account for 4% to 6% of the total attrition between waves, which was 5696 respondents, or 21.26% of Wave 1 respondents.

A.5

Appendix B: Mover vs Stayer Characteristics by Type of Attrition				
Unconditional Means: "Movers" vs "Stayers"	Refused		Non-Contact	
	(1)		(2)	
	Movers	Stayers	Movers	Stayers
Age	25	**31**	25	**28**
Years of Education	7.7	5.0	**7.9**	4.8
% Male	0.5	0.5	0.5	0.5
HH Income	9291	9924	**4542**	3127
HH Size	**5.1**	4.8	**4.8**	4.3
HH Children	**2.0**	1.6	1.9	1.9
HH Adults	3.1	3.1	**2.8**	2.4
Individual Income	2231	2400	1718	945
Labour Income (a)	2112	1745	987	1073
% Wage Employed	0.3	0.3	0.4	0.4
% Casual Employed	0.0	0.0	0.1	0.1
% Self Employed	0.1	**0.1**	0.0	0.0
% Unemployed	0.7	0.7	0.6	0.6
% Married	0.3	**0.4**	0.2	0.2
% Widowed	0.1	0.1	0.0	**0.1**
% Divorced	0.0	0.0	0.0	0.0
% Never Married	**0.5**	0.4	**0.6**	0.5
BMI	25.2	26.0	25.1	25.9
Perceived Health Index	3.9	3.8	**3.8**	3.6
% Overweight	0.3	0.2	0.2	0.2
% Obese	0.2	**0.2**	0.2	0.2
ADL Score	0.2	**0.4**	0.3	**0.5**
Depression Index	11.1	11.4	10.9	11.5
% Hypertension	0.2	**0.4**	0.3	**0.3**
N	392	1744	1990	724

Notes: (a) Labour Income is for employed respondents only.

Two sets of T-tests were run between movers and stayers. The first for those who were lost to refusal (column 1), and the second for those lost to non-contact (column 2). When means of movers and stayers differ at the 95% significance level, the larger of the two means is **Bold**.

Amongst refusors, stayers are older, and come from smaller households containing fewer children. They are also more likely to be self-employed and married. Movers are more likely to have never been married, and have lower obesity rates and cope better with daily activities, they are also less likely to suffer hypertension.

Among those lost to non-contact, movers are younger, more educated, come from larger and wealthier household and are less likely to have been widowed, or married at all. They record higher perceived health levels, fewer difficulties in daily activities, and lower incidence of hypertension. Many of these variables are obvious correlates of age, and the movers tend to be younger, more mobile and healthier.

A.6

Appendix C: Mover vs Stayer Characteristics by Attrition Status						
Unconditional Means: "Movers" vs "Stayers"	Attritors		Non-Attritors		Total	
	(4)		(5)		(6)	
	Movers	Stayers	Movers	Stayers	Movers	Stayers
Age	25	**30**	21	**27**	23	**27**
Years of Education	**6.4**	4.8	**5.3**	4.7	**6.6**	4.8
% Male	**0.5**	0.5	0.5	0.4	**0.5**	0.4
HH Income	5324	7930	4161	4098	**4820**	4533
HH Size	**4.8**	4.6	5.4	**5.8**	5.1	**5.6**
HH Children	**1.9**	1.7	2.3	**2.5**	2.1	**2.4**
HH Adults	2.9	**2.9**	3.1	**3.3**	3.0	**3.2**
Individual Income	1803	1973	978	933	1445	1051
Labour Income (a)	1175	**1548**	**1116**	684	**1150**	792
% Wage Employed	0.3	0.3	**0.3**	0.2	**0.3**	0.2
% Casual Employed	0.1	0.0	0.1	0.0	0.1	0.0
% Self Employed	0.0	**0.1**	0.0	**0.1**	0.0	**0.1**
% Unemployed	0.7	0.7	0.7	**0.8**	0.7	**0.8**
% Married	0.2	**0.3**	0.2	**0.3**	0.2	**0.3**
% Widowed	0.0	**0.1**	0.0	**0.1**	0.0	**0.1**
% Divorced	0.0	0.0	0.0	0.0	0.0	0.0
% Never Married	**0.6**	0.5	**0.7**	0.5	**0.6**	0.5
BMI	25.1	**25.9**	25.3	**26.1**	25.2	**26.0**
Perceived Health Index	3.8	3.7	**3.9**	3.5	**3.8**	3.5
% Overweight	0.2	0.2	0.2	0.2	0.2	0.2
% Obese	0.2	**0.2**	0.2	**0.2**	0.2	**0.2**
ADL Score	0.3	**0.4**	0.2	**0.5**	0.3	**0.4**
Depression Index	11.0	11.4	10.7	10.4	**10.9**	10.5
% Hypertension	0.3	**0.4**	0.2	**0.4**	0.2	**0.4**
N	2382	2468	1822	19276	4204	21744

Notes: (a) Labour Income is for employed respondents only.

Three sets of T-tests were run between movers and stayers: the first for all those who attrited (column 4), the second for those who remained in the panel (column 5), and the third on the total sample (column 6). When means of movers and stayers differ at the 95% significance level, the larger of the two means is **Bold**.

In all three groups, movers were younger, more educated, less likely to be married, to have ever been married or widowed, and less likely to be self-employed. They also recorded lower BMI's, had a lower incidence of obesity, coped better with daily activities, and had lower blood pressures.

Interesting patterns emerge regarding household composition and labour income. Among attritors, movers came from bigger households and reported lower labour incomes, but among those who remained in the panel, it is the stayers who come from larger households, and report lower labour incomes.

Mobility seems to be strongly age-related. There is little substantial evidence of upward mobility.

southern africa labour and development research unit

The Southern Africa Labour and Development Research Unit (SALDRU) conducts research directed at improving the well-being of South Africa's poor. It was established in 1975. Over the next two decades the unit's research played a central role in documenting the human costs of apartheid. Key projects from this period included the Farm Labour Conference (1976), the Economics of Health Care Conference (1978), and the Second Carnegie Enquiry into Poverty and Development in South Africa (1983-86). At the urging of the African National Congress, from 1992-1994 SALDRU and the World Bank coordinated the Project for Statistics on Living Standards and Development (PSLSD). This project provide baseline data for the implementation of post-apartheid socio-economic policies through South Africa's first non-racial national sample survey.

In the post-apartheid period, SALDRU has continued to gather data and conduct research directed at informing and assessing anti-poverty policy. In line with its historical contribution, SALDRU's researchers continue to conduct research detailing changing patterns of well-being in South Africa and assessing the impact of government policy on the poor. Current research work falls into the following research themes: post-apartheid poverty; employment and migration dynamics; family support structures in an era of rapid social change; public works and public infrastructure programmes, financial strategies of the poor; common property resources and the poor. Key survey projects include the Langeberg Integrated Family Survey (1999), the Khayelitsha/Mitchell's Plain Survey (2000), the ongoing Cape Area Panel Study (2001-) and the Financial Diaries Project.

SALDRU

www.saldru.uct.ac.za
Level 3, School of Economics Building, Middle Campus, University of Cape Town
Private Bag, Rondebosch 7701, Cape Town, South Africa
Tel: +27 (0)21 650 5696
Fax: +27 (0) 21 650 5797
Web: www.saldru.uct.ac.za

University of Cape Town · iYunivesithi yaseKapa · Universiteit van Kaapstad