

EXPLORING ATTRITION BIAS: THE CASE OF THE KHAYELITSHA PANEL STUDY (2000-2004)

JEREMY MAGRUDER* AND NICOLI NATTRASS**

Abstract

Attrition bias is a problem for users of panel data. Researchers need to know what socio-economic factors are associated with attrition, and whether this is of relevance for the kind of analysis they want to conduct. This paper discusses attrition bias in the 2000/2004 Khayelitsha panel study. It shows that women, shack-dwellers and people living in smaller households are more likely to attrit, but that the impact of these variables on the probability of attrition is relatively small. The implications for labour-market analysis are then explored using Mincerian earnings functions and a probit regression on whether respondents are wage-employed or not. The coefficients generated using a restricted sample of non-attriters do not differ significantly from those generated by the entire sample. This suggests that attrition bias in this particular data set is not a problem for this kind of labour market analysis.

JEL Classification: O12, O55

Keywords: attrition, Khayelitsha panel study, South Africa, Cape Town, attrition bias

1. INTRODUCTION

There are a growing number of panel studies in South Africa and other developing countries (see review in Lee, 2003). Panel studies are longitudinal surveys in which the same respondents are interviewed in successive 'waves'. Such studies thus have the potential to shed light on socio-economic dynamics such as changing labour-market participation, household composition, sources of income, demographic transitions *etc.* However, with each successive wave, panel studies inevitably suffer attrition as respondents die, refuse to participate or cannot be tracked to a new location. If this is not a random process, then the value of the panel data set may be undermined significantly by 'attrition bias'.

Two key issues arise for data users with regard to attrition bias. The first is, does it exist – and if so, which categories of respondents are disproportionately affected? The second question is, given the existence of attrition bias, does it matter for the specific issue being analysed? Depending on the nature of the attrition bias, it will matter more for some analytical questions than others. The accepted wisdom is that researchers using panel data should be cognizant of both issues before using panel data in their modelling.

This short paper examines attrition bias in the two-wave Khayelitsha panel study (conducted in 2000 and 2004). We start off with exploratory probit analysis to explore which socio-economic characteristics are associated with attrition. The analysis suggests that women, shack-dwellers and people living in smaller households are more likely to attrit, but that the impact of these variables on the probability of attrition is relatively

* Yale University.

** University of Cape Town. The authors are thankful to anonymous referees for comments made on an earlier draft.

useful information about changes in labour-market participation *etc* – but only if we are reasonably confident that attrition bias is not a problem for labour-market analysis.

2. REASONS FOR ATTRITION

Table 1 provides information about the reasons provided by fieldworkers as to why the missing respondents were not interviewed.

Only 9 per cent of the non-responses were outright refusals. The biggest single problem was not being able to locate an entire household (largely because the address information was poor, or because, in two cases, entire enumerator areas had disappeared as a result of construction work in the area. Other problems arise in the process of the fieldwork itself, such as information from fieldworkers being incomprehensible or entirely absent with regard to why respondents were not re-interviewed.

Table 2 lists attrition for men and women in different potential data (sub) sets. The 'four-year Residents' sample excluded those who had died or had moved out of the area. Depending on the sample, attrition is either 41.1% or 29.5%. Such attrition rates are broadly comparable with other developing country panel studies such as the India

Table 1. Reasons for Attrition between 2000 and 2004

Reason Given for Failure to Interview the Individual	N	%
Lived in the enumerator area which could not be found (ca = 1,066,722) owing to a poor initial aerial photograph and construction in the area	20	5.4
Enumerator area found, but household could not be found	109	29.1
Individuals unknown by those household members who could be located	48	12.8
Not available for interview (even after several visits)	10	2.7
Refused	9	2.4
Deceased	35	9.4
Moved within Cape Town (no address given)	33	8.8
Moved outside Cape Town	61	16.3
Moved, location unknown	32	8.6
Incarcerated	1	0.3
On holiday	1	0.3
Wrong person identified and surveyed	15	4.0
Total	374	100.0
Consent forms not signed (3) or missing observation for consent (4)	7	
Interviews conducted	570	
Total (for which we have information) (570 + 373)	944	
Potential respondents unaccounted for in the re-survey (survey management problems)	22	
Total potential sample	966	

Table 2. Attrition in the Khayelitsha Survey

	Total Sample	Four-year Residents
Individual, total	41.1%	29.5%
Individual, women	37.6%	25.0%
Individual, men	46.1%	35.2%
Individual Sample Size	966	804
Household	32.7%	24.9%
Household Sample Size	434	365
Individuals in found households	22.0%	11.8%
Women in found households	19.5%	9.6%
Men in found households	25.5%	14.9%
Sample Size	727	643

Additional Rural Incomes Survey (33% attrition), the Bolivia Integrated Child Development Program panel survey (35% attrition), the Malaysian Family Life Survey (27% attrition) and the Kenyan Ideational Change Survey (32% attrition) – see Lee (2003). For the analysis of attrition in this paper, we use the overall sample, which has the highest attrition rate and therefore is likely to manifest the greatest attrition bias.

3. CHARACTERISTICS OF ATTRITORS

Table 3 provides mean estimates (and a 95% confidence interval around those mean estimates) for key socio-economic characteristics by three different types of attritor: those who we know moved away; those who lived in ‘lost’ households (*i.e.* household that could not be found at all by the field-workers) and those who attrited for other reasons. Unemployment in Table 3 (and in the rest of this analysis) follows the ‘broad’ unemployment definition: *i.e.* it includes all those aged 18 or above who were without work, and who said that they wanted to work (see Natrass, 2002 and Natrass and Walker, 2005 for a discussion of labour-market categories used in the KMP survey).

Table 3 shows that there were no statistically significant differences in mean age, the proportion who were men, mean household income or mean individual income between the different types of attritors (because the 95% confidence interval overlaps in all cases). There were also no statistically significant differences in the percentage of people in different labour-market states. The only statistically significant differences were that those who lived in lost households were more likely to have lived in smaller households than other types of attritors, that mean years of education were lower for living in lost households than those who moved away, and that those who moved away were more likely to have been living in shacks than other attritors.

In other words, there are relatively few observable differences between the different kinds of attritors. It thus makes sense to analyse them as a single group. (Nevertheless, in order to check whether attritors who were known to have moved away had noticeably different characteristics, Table 7 includes a regression model of determinants of attrition for this sub-group of attritors as well as for all attritors).

Table 4 presents descriptive statistics for attritors and non-attritors including t-tests of differences between the mean values for selected variables. It shows that for both men and women, attritors are more likely to come from smaller households and live in shacks – but that otherwise, different patterns are evident for men and women. Younger men are more likely to attrit (age is not a significant difference for women), and women with lower individual incomes and living in lower income households are more likely to attrit (income is not significant for men). Unemployed females are also more likely to attrit. Households which attrit (*i.e.* those households which could not be found at all), tend to be poorer and to be found in shack settlements.

4. MULTIVARIATE ANALYSIS OF ATTRITION

To check for the statistical significance of these differences in a multivariate analysis, we can run probits on the probability of attriting. Model 1 in Table 5, which uses the entire sample, shows that males, shack-dwellers, younger people and those living in small and low-income households have a significantly greater probability of attriting than other

Table 3. Selected Characteristics (from KMP 2000 Survey) of Attritor by Type of Attrition

	Moved away		Lost household		Other attrition		Total attrition		Non-attritors	
	Mean	95% ci	Mean	95% ci	Mean	95% ci	Mean	95% ci	Mean	95% ci
Monthly household income (R)	2,090	1,461-2,720	2,034	1,564-2,501	2,928	2,056-3,800	2,377	1,975-2,779	3,315	2,524-4,106
Individual income (R)	591	430-755	681	474-887	687	499-876	655	547-762	688	598-779
Household size	4.4	4.0-4.8	3.5	3.1-3.8	4.9	4.4-5.3	4.3	4.0-4.9	5.1	4.9-5.3
Age	30.9	28.9-32.8	34.0	32.2-35.8	34.1	32.0-36.2	33.0	31.9-34.1	34.8	33.7-35.8
Education	9.3	8.8-9.8	8.0	7.4-8.6	8.5	8.0-9.0	8.6	8.3-8.9	8.4	8.2-8.7
% Male	43.6%	34.9-52.4%	53.5%	44.8-62.2%	47.9%	39.5-56.2%	48.4%	43.4-53.3%	40.3%	36.3-44.3%
Wage-employed	28.6%	20.6-36.6%	31.8%	23.6-40.0%	33.6%	25.7-41.5%	31.4%	26.8-36.0%	34.9%	30.9-38.9%
Casual employed	2.4%	0-5.1%	3.1%	0.1-6.1%	1.4%	0-3.4%	2.3%	0.8-3.8%	2.8%	1.4-4.2%
Self employed	4.0%	0.5-7.4%	9.3%	4.2-14.4%	5.0%	1.4-8.7%	6.1%	3.7-8.4%	8.1%	5.8-10.3%
Unemployed	58.8%	50.0-67.8%	44.0%	34.8-53.1%	49.2%	40.2-58.2%	50.1%	45.9-55.9%	50.7%	45.5-55.9%
Lives in a shack	74.6%	66.9-82.3%	72.9%	65.1-80.4%	57.9%	49.6-66.1%	68.1%	63.5-72.7%	55.5%	51.4-59.6%
N	126		129		140		395		570	

Table 4. Selected Characteristcs of Attritor (from KMP 2000 Survey)

Statistics	Monthly household income	Individual income	Household size	Age	Education	Wage-employed	Causal employed	Self-employed	Unemployed	Lives in a shack
Male, non-attritors										
Mean	R1,766	R845	4.96	35.9	7.8	0.475	0.036	0.054	0.358	0.565
Standard deviation	R1,476	R1,095	2.40	13.0	3.5	0.501	0.186	0.226	0.481	0.497
Male Attritors										
Mean	R1,619	R886	4.14	33.2	8.0	0.417	0.037	0.054	0.422	0.691
Standard deviation	R1,541	R1,311	2.31	11.2	3.6	0.494	0.190	0.000	0.495	0.463
Difference in means for male non-attritors and male attritors										
Difference	R148	-R41	0.82***	2.7*	-0.2	0.058	-0.001	0	-0.06	-0.126**
t-test	0.998	-0.344	3.557	2.303	-0.663	1.181	-0.083	0.01	-1.31	-2.686
Female non-attritors										
Mean	R1,848	R1,813	5.25	34.0	8.8	0.29	0.0240	0.123	0.464	0.548
Standard deviation	R1,690	R8,103	2.43	11.9	3.2	0.46	0.1531	0.329	0.450	0.498
Female Attritors										
Mean	R1,329	R583	4.35	32.9	8.4	0.25	0.0099	0.095	0.552	0.672
Standard deviation	R1,230	R1,098	2.46	11.5	3.9	0.44	0.0995	0.293	0.500	0.471
Difference in means for female non-attritors and female attritors										
Difference	R520***	R144*	0.90***	1.1	0.4	0.04	0.014	0.028	-0.09**	-0.123***
t-test	4.135	1.804	4.144	1.058	1.325	1.00	1.28	1.03	-1.979	-2.891
Statistics	Monthly household income (non attritors)	Monthly household income (attritors)	Monthly household income (non-attritors)	Household size	Household size (non-attritors)	Household size (attritors)	Live in a shack	Live in a shack (non attritors)	Live in a shack (attritors)	
Mean	R1,664	R1,364	R1,364	4.43	4.43	3.41	0.564	0.564	0.665	
Standard deviation	R1,522	R1,458	R1,458	2.17	2.17	1.97	0.497	0.497	0.473	
Difference			R300**			1.02***			-0.100**	
t-test			2.005			4.9			-2.06	

The t-test is a two-sample t-test with unequal variances. * indicates significance at the 10% level and ** at the 5% level. *** at the 1% level. The labour force categories are based on 'lmstatus2' as described in Natrass (2002).

Table 5. Probit regressions on Attrition

Dependent Variable: Attritor	1	2	3	4	5	6	7
Age	Entire sample	Women only	Men only	Live in a shack	Live in a brick house	Men living in a shack	Women living in a shack
dF/dx	** -0.0034	-0.0027	-0.0040	-0.0029	* -0.0043	* -0.007	0.0023
z	-2.07 (0.038)	-1.19 (0.235)	-1.66 (0.097)	-1.27 (0.204)	-1.85 (0.064)	-2.23 (0.026)	0.70 (0.484)
Female	**			*			
dF/dx	-0.0757	-0.0850		-0.0850	-0.0682		
z	-2.24 (0.025)	-1.93 (0.054)		-1.93 (0.054)	-1.34 (0.180)		
P > z							
Education							
dF/dx	-0.0053	-0.0101	0.0019	-0.0039	-0.0101	-0.0054	0.0013
z	-0.96 (0.337)	-1.39 (0.165)	0.23 (0.820)	-0.56 (0.576)	-1.12 (0.262)	-0.50 (0.614)	0.14 (0.892)
P > z	*	**					**
Total household income							
dF/dx	-0.00002	-0.00003	-0.00000	-0.00003	-0.00001	-0.00000	-0.00006
z	-1.68 (0.093)	-2.03 (0.020)	-0.17 (0.865)	-1.59 (0.112)	-0.70 (0.482)	-0.25 (0.806)	-2.18 (0.029)
P > z							
Working							
dF/dx	-0.048	-0.0787	0.049	-0.0865	0.0044	-0.0108	-0.1715
z	-0.76 (0.447)	-0.98 (0.327)	0.05 (0.961)	-1.00 (0.319)	0.05 (0.960)	-0.09 (0.932)	-1.42 (0.157)
P > z							
Unemployed							
dF/dx	0.0059	-0.0077	0.0506	0.0268	-0.0369	0.0699	-0.0127
z	0.09 (0.926)	-0.10 (0.923)	0.49 (0.625)	0.31 (0.755)	-0.41 (0.684)	0.55 (0.581)	-0.11 (0.915)
P > z	***	**	**				
Living in a shack							
dF/dx	0.1147	0.1058	0.1254				
z	3.31 (0.001)	2.32 (0.020)	2.34 (0.019)				
P > z	***	***	***	***	***	***	***
Household size							
dF/dx	-0.0367	-0.0373	-0.0360	-0.0349	-0.0392	-0.0380	-0.0319
z	-5.21 (0.000)	-4.06 (0.000)	-3.32 (0.001)	-4.08 (0.000)	-3.20 (0.001)	-2.90 (0.004)	-2.80 (0.005)
P > z							
Number of observations	929	528	401	562	367	250	312
Pseudo R-squared	0.0506	0.0577	0.0419	0.0459	0.0374	0.0506	0.0537

* indicates significance at the 10% level and ** at the 5% level. *** at the 1% level.

Table 6. Probit Estimates

Log likelihood = -261.01283					Number of observations = 401	
					Prob > chi2 = 0.0000	
					Pseudo R2 = 0.0507	
	Coefficient	Standard Error	z-score	P > z	Confidence Interval	
Household Attritor	-0.14279	0.0318613	-4.48	0.000	-0.2052369	-0.090343
Household size	-0.0000357	0.0000415	-0.86	0.390	-0.000117	0.0000457
Household income	0.2403315	0.1341033	1.79	0.073	-0.0225062	0.5031692
Shack/house	0.3117429	0.1701663	1.83	0.067	-0.0217769	0.6452626
Constant						

respondents. The Table reports marginal effects – so for example, in model 1, we can conclude that being female reduces the probability of attriting by 7.6 percentage points (controlling for all other variables held constant at their mean values). Likewise, the probability of attriting is reduced by 3.7 percentage points if household size increases by one member. Note that household income is significant only at the 10% level and the size effect is very small (an increase in household income of R100 only reduces the probability of attriting by 0.2 percentage points).

Models 2 and 3 run the regression conditional on gender, and models 4 and 5 run the regression conditional on whether the respondent lives in a shack or a house. The results show that once other factors are controlled for, the impact of household income loses significance, whereas the effect of household size remains important in all 4 models. Model 6 runs the regression selecting only male respondents living in shacks.

Apart from household size, age is the only (marginally significant) determinant of attrition. For women, however (model 7), those living in smaller and poorer households are more likely to attrit. However in all cases where the independent variables are statistically significant, the size impact (on the probability of attrition) is relatively small.

We may also be concerned about the effect of household size and whether the household lives in a shack or not on household attrition. The probit regression reported in Table 6 shows that small households are more likely to attrit – as are households living in shacks – but that household income is not a significant predictor of household attrition. Note that this regression is run at the household level, with the dependent variable taking a value of 0 if the household was found, and 1 if it was not.

Given that household size is a strong predictor for household attrition, this suggests that either smaller households are more mobile or systematically different even conditional on age, education, etc., or the fieldworkers had more difficulty tracking people from smaller households (as there would have been fewer individual points of reference to help with the search for respondents). As is clear from Table 5, household size is also a strong predictor for individual attrition.

Table 3 showed that this was particularly a problem for those who attrited because the entire household could not be found. Unsurprisingly, then, Table 7 shows that conditional on the household being found, household size has no significant effect on individual attrition probabilities.

Model 5 in Table 7 explores the determinants of attrition when attritors are defined as those who were known to have moved away. It shows that those education and living in a shack were significant predictors of attrition, but that the impact was small: for every extra year of education, the probability of attrition increased by only 1.9 percentage points, and living in a shack increased the probability of attriting by 10 percentage points.

Table 7. Probit regressions on Attrition

Dependent Variable: Attritor	1	2	3	4	5	6
	Entire sample	Found households only	Women in found households only	Men in found households	Only attritors who moved away	Entire sample
Age	**	-0.00229	0.0004	**	-0.0018	-0.0018
dF/dx	-2.07	-1.44	0.02	-2.13	-0.73	-1.05
z	(0.038)	(0.149)	(0.982)	(0.033)	(0.468)	(0.295)
P > z	**	**				**
Female	-0.0757	-0.07006			0.0196	-0.0784
dF/dx	-2.24	-2.15			0.40	-2.31
z	(0.025)	(0.031)			(0.69)	(0.021)
P > z					**	
Education	-0.0053	-0.0009	-0.0021	-0.0011	0.0190	-0.0060
dF/dx	-0.96	-0.17	-0.29	-0.12	2.4	-1.08
z	(0.337)	(0.337)	(0.771)	(0.906)	(0.017)	(0.280)
P > z	*					*
Total household income	-0.00002*	-0.00000	-0.00002	0.00002	-0.00000	-0.00002
dF/dx	-1.68	-0.44	-1.14	0.90	-0.30	-1.84
z	(0.093)	(0.662)	(0.252)	(0.370)	(0.978)	(0.065)
P > z						
Working	-0.048	-0.048	-0.383	-0.0571	0.0429	-0.0095
dF/dx	-0.76	-0.82	-0.54	-0.57	0.43	-0.15
z	(0.447)	(0.412)	(0.592)	(0.566)	(0.670)	(0.883)
P > z						
Unemployed	0.0059	-0.0477	0.0084	0.0016	0.1305	0.0311
dF/dx	0.09	-0.03	0.11	0.02	1.34	0.49
z	(0.926)	(0.974)	(0.908)	(0.987)	(0.180)	(0.626)
P > z	***				*	***
Living in a shack	0.1147	0.0537	0.0712	0.0281	0.1011	0.1132
dF/dx	3.31	1.64	1.72	0.53	1.9	3.25
z	(0.001)	(0.101)	(0.085)	(0.597)	(0.057)	(0.001)
P > z	***					***
Household size	-0.0367	0.00926	0.0063	0.0134	0.0052	-0.0337
dF/dx	-5.21	1.45	0.79	1.26	0.52	-4.76
z	(0.000)	(0.147)	(0.427)	(0.208)	(0.606)	(0.000)
P > z						***
Recent resident						0.1602
dF/dx						2.68
z						(0.007)
P > z						
Recent migrant						0.0354
dF/dx						0.51
z						(0.607)
P > z						929
Number of observations	929	702	409	293	384	929
Pseudo R-squared	0.0506	0.0238	0.0243	0.0351	0.0437	0.0652

* indicates significance at the 10% level and ** at the 5% level. *** at the 1% level.

One might expect attrition to be higher amongst those who were recent migrants to Cape Town, and/or had moved recently to the dwelling they were living in at the time of the 2000 KMP survey. Model 6 in Table 7 includes two further dummy variables: 'recent migrant' (those who had moved to Cape Town since 1998) and 'recent resident' (those who had moved to the household since 1998).

In short, the probit regression models reported thus far suggest that household size, gender and housing type are significant determinants of attrition, but their impact on the probability of attrition is relatively small. The sample is thus not strongly biased as far as these observables are concerned. The same cannot be said for those who had moved recently to the household – as their chances of attrition are high. If one were interested in core household members rather than all individuals – some of whom would have joined the household as a short-term accommodation strategy with the intention of moving on as soon as they found a job or suitable alternative accommodation – then this form of attrition bias may not be so important. However, if one were interested in tracking individual transitions, then this form of attrition bias needs to be borne in mind.

The extent to which the sample bias matters is of course dependent on the research question being posed. As Cichello observes, the existing literature on the problem of panel attrition generally concludes that despite mean differences in the characteristics of attritors, structural relationships are not necessarily affected (despite the non random nature of attrition) – but that this should be tested by researchers on a case-by-case basis as to whether the relationships matter because no general conclusions can be made – even within the same data set (Cichello, 2001:16; see also Alderman *et al.*, 2001). For example, if the main objective is to conduct labour-market analysis – and if factors such as household size and housing type are incidental to it – then the bias reported here for the Khayelitsha survey is probably incidental. This is explored below.

5. SOME EXPLORATORY LABOUR-MARKET ANALYSIS

This final section of the paper provides some exploratory labour-market analysis to see whether attrition bias matters. In this regard, we follow Falaris (2003) in trying to estimate classical equations to tell if coefficient estimates appear biased by using the sample of non-attritors *versus* the overall sample. Two obvious equations are a Mincerian earnings equation and a probit on whether a person has wage employment (as reported in Table 8).

In none of these simple analyses was the slope or intercept coefficients of non-attritors individually or jointly significantly different from the sample at large (shown in Wald and Chi2 tests reported at bottom of Table 8). In other words, for this particular analysis, the coefficients generated using a restricted sample of non-attritors do not differ significantly from those generated by the entire sample.¹

Now let us examine labour force participation. Table 9 summarises the results for labour-force participation using the full sample, the sample of attritors only, and the sample of non-attritors. It shows that the distribution of labour-force categories is close to that for the full sample as a whole and that there is no significant difference in the

¹ This same conclusion can be drawn after running the same regressions on the full sample but this time defining attrition only for those who had been known to have moved away.

Table 8. Earnings Function and Probability of being in Wage Employment

	Dependent variable: log monthly wage		Dependent variable: log monthly wage		Dependent variable: probability of being in wage employment		Dependent variable: probability of being in wage employment	
	Men and Women	Women	Men	Men and Women	Women	Men		
Education coefficient	0.066	0.0941	0.0396	0.0646	0.1068	0.0136		
t(t)-statistic	4.40	3.64	2.21	3.31	3.90	4.98		
Age coefficient	0.0232	-0.0368	0.086	0.1770	0.1378	0.2339		
t-statistic	0.80	-0.84	2.30	6.03	3.58	4.98		
Age squared coefficient	-0.002	0.001	-0.001	-0.0003	-0.0012	-0.0026		
t-statistic	-0.55	1.23	-2.45	0.38	2.49	0.44		
Female	***			***				
coefficient	-0.4521			-0.4800				
t-statistic	-5.88			5.32				
Education*attrit	0.0135	0.0254	0.0239	0.0110	-0.0030	0.0378		
t-statistic	0.59	0.21	1.31	0.38	-0.07	0.92		
Age*attrit	-0.0150	-0.0356	-0.0441	-0.0022	0.0779	-0.0877		
t-statistic	-0.35	-0.33	-0.89	-0.05	1.06	-1.34		
Age squared*attrit	0.0002	0.0004	0.0007	-0.0002	-0.0012	-0.0877		
t-statistic	0.45	0.28	1.31	-0.45	-1.38	-1.34		
Attritor								
coefficient	0.0761	0.4042	0.3777	0.2532	-0.9701	1.395		
t-statistic	0.09	0.21	0.36	0.27	-0.67	1.05		
Adj/Pseudo R squared	0.1581	0.1491	0.1270	0.1101	0.1149	0.0908		
N	291	132	159	950	538	412		
Wald/Chi2 test	F(4,282) = 0.16 Prob > F = 0.958	F(4,124) = 0.23 Prob > F = 0.921	F(4,151) = 1.5 Prob > F = 0.206	Chi2(4) = 9.31 Prob > Chi2 = 0.054	Chi2(4) = 4.59 Prob > Chi2 = 0.332	Chi2(4) = 6.48 Prob > Chi2 = 0.145		

Wald / Chi2 tests that the coefficients on attritor, age*attritor, age squared*attritor and education*attritor = 0

Table 9. Labour-Force Participation in 2000

	Full sample in 2000	Non-attriters	Attriters
Basic Labour-Force Categories			
Employed	46.6%	49.0%	43.0%
Active job seekers	19.8%	18.5%	21.7%
Those wanting work but not seeking it	25.2%	23.7%	27.3%
Non-labour force participants	8.5%	8.8%	8.0%
	100%	100%	100%
N	945	557	388
	Pearson chi2 (3) = 4.2705 Pr = 0.234		
Expanded Labour-Force Categories (Natrass, 2002)*			
Wage employed	35.2%	36.6%	33.3%
Self employed	8.7%	9.5%	7.5%
Casually employed	2.7%	2.9%	2.3%
Searching unemployed	19.8%	18.5%	21.7%
Network searching unemployed	6.8%	7.4%	5.9%
Marginalised unemployed	18.4%	16.3%	21.4%
Non-labour force participants	8.5%	8.8%	8.0%
	100%	100%	100%
N	945	557	388
	Pearson chi2 (6) = 7.2972 Pr = 0.294		

* Following Natrass (2002), the searching unemployed are those engaging in active job search; the network searching are those who are not actively seeking work but are relying on their social networks to find them work, and the marginalised unemployed are those who are neither searching for work nor have any social contacts who can do it for them (see also Natrass and Walker, 2005).

distributions within different samples. In other words, at least as far as *prior* labour-force participation is concerned, attrition did not bias the sample to any significant extent.

6. CONCLUSION

To what extent did attrition between the 2000 and 2004 waves of the Khayelitsha panel survey bias the sample? The short answer is that attrition bias is evident (but for the most part not particularly strong) and is likely to be a problem only for particular research questions (*e.g.* those affected by higher attrition rates among new household members). For general labour-market enquiries, the attrition bias does not seem significant – although researchers should remain alert to the possibility that attrition bias may be influencing their results.

REFERENCES

- ALDERMAN, H., BEHRMAN, J., KOHLER, H., MALUCCIO, J. and COTTS-WATKINS, S. (2001). "Attrition in Longitudinal Household Survey Data: Some Tests for Three Developing Country Samples", in *Demographic Research*, 5(4). Available on www.demographic-research.org
- CICHELO, P. (2001). "Did African Workers in Post-Apartheid KwaZulu-Natal Really Get Ahead? Robustness Checks in the Light of Panel Attrition?" draft paper, January 2001.
- COETZEE, C. (2005). The Impact of Highly Active Antiretroviral Therapy (HAART) on Perceived Health and Labour Supply in Khayelitsha. Master's Thesis, School of Economics, University of Cape Town.
- COETZEE, C. and NATRASS, N. (2004). Living on AIDS Treatment: A Socio-Economic Profile of Africans Receiving Antiretroviral Therapy in Khayelitsha, Cape Town. *CSSR Working Paper* No. 71, Cape Town, Centre for Social Science Research, University of Cape Town. Available on www.cssr.uct.ac.za
- CRANKSHAW, O., WELCH, M. and BUTCHER, S. (2001). "GIS Technology and Survey Sampling Methods: The Khayelitsha/Mitchell's Plain 2000 Survey" in *Social Dynamics*, 27(2): 156-174.
- FALARIS, E. (2003). "The Effect of Survey Attrition in Longitudinal Surveys: Evidence from Peru, Cote d'Ivoire, and Vietnam", in *Journal of Development Economics*, vol. 70: 133-157.
- LEE, U. (2003). "Panel Attrition in Survey Data: A Literature Review." *CSSR Working Paper* No. 41, Cape Town, Centre for Social Science Research, University of Cape Town. Available on www.cssr.uct.ac.za

- MAGRUDER, J. and NICOLI NATTRASS. (2005). "Attrition in the Khayelitsha Panel Study (2000-2004)". *CSSR Working Paper* No. 123, Cape Town, Centre for Social Science Research, University of Cape Town. Available on www.cssr.uct.ac.za
- NATTRASS, N. (2002). "Unemployment, employment and Labour Force Participation in Khayelitsha/Mitchells Plain", *CSSR Working Paper* No. 12, University of Cape Town. Available on www.cssr.uct.ac.za
- NATTRASS, N. and WALKER, R. (2005). "Unemployment and Reservation Wages in Working-Class Cape Town", in *South African Journal of Economics*, vol.73, no.3.