

Data Note

Pakistan Panel Household Survey: Sample Size and Attrition

DURR-E-NAYAB and G. M. ARIF

1. INTRODUCTION AND BACKGROUND

The socio-economic databases in Pakistan, as in most countries, can be classified into three broad categories, namely registration-based statistics, data produced by different population censuses and household survey-based data. The registration system of births and deaths in Pakistan has historically been inadequate [Afzal and Ahmed (1974)] and the population censuses have not been carried out regularly. The household surveys such as Pakistan Demographic Survey (PDS), Labour Force Survey (LFS) and Household Income Expenditure Survey (HIES) have been periodically conducted since the 1960s. These surveys have filled the data gaps created by the weak registration system and the irregularity in conducting censuses. The data generated by the household surveys have also enabled social scientists to examine a wide range of issues, including natural increase in population, education, employment, poverty, health, nutrition, and housing. All these surveys are, however, cross-sectional in nature so it is not possible to gauge the dynamics of these social and economic processes, for example the transition from school to labour market, movement into or out of poverty, movement of labour from one state of employment to another. A proper understanding of such dynamics requires longitudinal or panel datasets where the same households are visited over time. Since panel surveys are complex and expensive to carry out, they are not as commonly conducted as the cross-sectional surveys anywhere in the world and in Pakistan they are even rarer.

One of the available panel surveys in Pakistan has been conducted by International Food Policy Research Institute (IFPRI) over a period of five years from 1986 to 1991 covering 800 households. The IFPRI sample comprised rural areas of only four districts with no representation from Balochistan and urban areas of the country. In these five years the sampled households were almost visited biannually. Another two-round panel data available in the country is that of the Pakistan Socio-Economic Survey (PSES) carried out by the Pakistan Institute of Development Economics (PIDE) in 1998-99 and

Durr-e-Nayab <dnayab@gmail.com> is Chief of Research at the Pakistan Institute of Development Economics, Islamabad. G. M. Arif <gmarif@pide.org.pk> is Joint Director at the Pakistan Institute of Development Economics, Islamabad.

Authors' Note: The authors are thankful to Shujaat Farooq for his help in the analysis regarding attrition of the sample. Thanks are due to Syed Majid Ali and Saman Nazir as well for their help in the tabulation for this paper. Usual disclaimer applies.

2001 in the rural as well as urban areas of Pakistan. Both the IFPRI and the PSES panels could not be continued after the above-mentioned rounds.

In 2001, the PIDE took a major initiative, with the financial assistance of the World Bank, to revisit the IFPRI panel households after a gap of 10 years. The sample was expanded from four to 16 districts, adding districts from all four provinces. Continuing to be a rural survey, it was named the Pakistan Rural Household Survey (PRHS). The second round of the PRHS was carried out in 2004 while the third round was completed in 2010. The third round marked the addition of the urban sample to the existing survey design of the PRHS, as a result—the Survey was named as the Pakistan Panel Household Survey (PPHS).

Attrition bias can affect the findings of the subsequent rounds of a panel survey, so it is important to examine the extent of sample attrition and determine whether it is random or has affected the representativeness of the panel sample. After conducting three rounds of the PRHS-PPHS there is a need to evaluate the panel dataset for attrition bias. The present paper looks into the socio-demographic profile of the sample over the three rounds and evaluates the presence, or otherwise, of an attrition bias. The paper, thus, has three major objectives, which are to:

- (a) Describe the sample size of three rounds of the panel survey
- (b) Analyse the extent of sample attrition and analyse whether it is random, and
- (c) Examine the socio-demographic dynamics of household covered in three rounds.

2. SELECTION OF DISTRICTS AND PRIMARY SAMPLING UNITS (PSUs)

As noted earlier, the IFPRI panel (1986-1991) was limited to the rural areas of four districts, namely Dir in Khyber Pakhtunkhwa (KP), Attock and Faisalabad in Punjab and Badin in Sindh. A rural sample based on these districts cannot be considered representative of the rural areas spread across more than 100 districts of the country. To give more representation to the uncovered areas 12 new districts were added to the PRHS-I round carried out in 2001. From KP two new districts, Mardan and Lakki Marwat, were added to give representation to the Peshawar-Mardan valley and the Kohat-Dera Ismail Khan belt, respectively. The Hazara belt of KP still needs to be added for an even better representation. Three districts from south Punjab (Bahawalpur, Vehari and Muzaffargarh) and one district from central Punjab (Hafizabad) were also included in the PRHS-I. By this addition, all the three broad regions of Punjab, north, central and south, have their representation in the panel survey (Table 1). The three added districts from Sindh were Mirpurkhas, Nawabshah and Larkana. Balochistan was not part of the IFPRI panel so the PRHS included three districts from Balochistan, namely Loralai, Khuzdar and Gawadar (Table 1).

For the rural sample a village or deh is considered as the PSU. Table 1 presents the number of rural PSUs by district. It is noteworthy that there were 43 PSUs (or village/deh) in four districts of the IFPRI panel (Attock, Dir, Badin and Faisalabad). From the 12 new districts, PRHS selected 98 more PSUs (villages/deh) randomly. The total rural PSUs, after all the additions and inclusions, now stand at 141 as can be seen in Table 1. For details regarding each selected PSU, their respective tehsils, districts and provinces see Table A1, A2, A3 and A4 in the Annexure.

Table 1
Primary Sampling Units (PSUs) by Province and District

Province	Districts	Number of PSUs	
		Rural	Urban ^c
Punjab	Faisalabad ^a	6	16
	Attock ^a	7	4
	Hafizabad ^b	10	4
	Vehari ^b	10	4
	Muzaffargarh ^b	9	4
	Bahawalpur ^b	9	7
Sindh	Badin ^a	19	3
	Nawab Shah ^b	8	4
	Mirpur Khas ^b	8	4
	Larkana ^b	11	7
KP	Dir ^a	11	2
	Mardan ^b	7	6
	Lakki Marwat ^b	5	2
Balochistan	Loralai ^b	7	2
	Khuzdar ^b	7	3
	Gwadar ^b	7	3
	Total	141	75

Note: PRHS-I (2001) and PPHS (2010) covered all districts. PRHS-II (2004) was limited to 10 districts of Punjab and Sindh.

a. Districts included in the IFPRI panel.

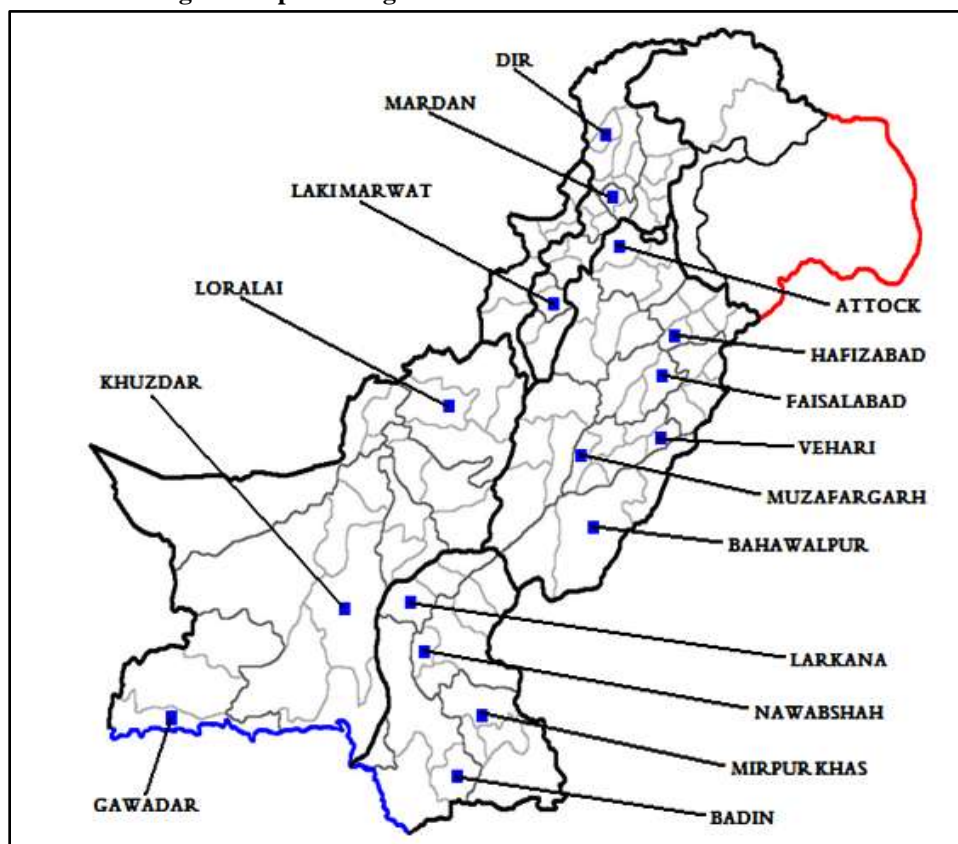
b. New districts added since 2001.

c. Included only in PPHS-2010.

It is worth mentioning here that the second round of the panel survey, PRHS-II, was carried out only in the rural areas of Punjab and Sindh. Because of security concerns the other two provinces, KP and Balochistan, could not be covered in this round.

The urban sample was added in the third round (PPHS) carried out in 2010 in all 16 districts. A selected district was the stratum for the urban sample. All the urban localities in each district were divided into enumeration blocks, consisting of 200 to 250 households in each block. In total, 75 urban enumeration blocks (PSUs) were selected randomly for the third round (PPHS-2010).

The scatter of the selected districts, as can be seen from Figure 1, is a good indicator of the geographical coverage of the districts covered under the PPHS. The sample covers the whole of the country, strengthening its representativeness.

Fig. 1. Map Showing Selected Districts for the PPHS-2010

3. HANDLING THE SPLIT HOUSEHOLDS

Before discussing the sample size, it is important to understand how the split households have been dealt with in the panel survey. A split household is defined as a new household where at least one member of an original panel household has moved in and is living permanently. This movement of a member from a panel household to a new household could be due to his/her decision to live separately with his/her family or due to marriage of a female member. If split households are not handled properly, the demographic composition of the sampled households is likely to change over time.

In the rounds two and three of the PRHS-PPHS split households were also interviewed. They, however, were only those households that were residing in the same village as the original panel household. In other words, movement of panel households or their members residing out of the sampled villages were not followed because of the high costs involved in this type of follow-up.

4. SAMPLE SIZE OVER THE DIFFERENT ROUNDS

The size of the sample for each round of the panel survey is shown in Table 2. The total size varies from 2721 households in 2001 to 4142 households in 2010. These variations, as discussed earlier, are for three reasons. First, the PRHS-II carried out in 2004 was limited to

two provinces, Punjab and Sindh, while the other two rounds covered all four provinces. Second, in the PRHS-II as well as the PPHS-2010, split households were also interviewed (Table 2). Third, urban sample was added in the third round, PPHS, 2010.

As can be seen from Table 2, in the PRHS-I, carried out in 2001, the total sample consisted of 2721 rural households. The sample size decreased to 1614 households in PRHS-II (2004) because of the non-coverage of two provinces. However, 293 split households were interviewed in PRHS-II to raise the total sample size to 1907 households. Table 2 shows that in the PPHS-2010 the total rural households interviewed in four provinces were 2800, out of which 2198 were panel households and the remaining 602 were split households. With the addition of 1342 urban households, the total sample size of the PPHS 2010 accounted for a total of 4142 households (Table 2).

Table 2

Households Covered during the Three Waves of the Panel Survey

	PRHS-I 2001	PRHS-II 2004			PPHS-2010				
		Panel House- holds	Split House- holds	Total	Panel House- holds	Split House- holds	Total Rural house- holds	Urban House- holds	Total Sample
Pakistan	2721	1614	293	1907	2198	602	2800	1342	4142
Punjab	1071	933	146	1079	893	328	1221	657	1878
Sindh	808	681	147	828	663	189	852	359	1211
KP	447	–	–	–	377	58	435	166	601
Balochistan	395	–	–	–	265	27	292	160	452

Source: PRHS 2001, 2004 and PPHS 2010 micro-datasets.

Four features of the three rounds of the panel data are noteworthy, which are as follows:

- (i) Urban households, which have been included for the first time in the sample in the third round (PPHS) held in 2010, are not panel households. Essentially, the urban sample can be analysed as a cross-sectional dataset at present and after their coverage in the next round of the survey they can be treated as panel households.
- (ii) Split households are not strictly panel households, particularly those where a female has moved due to her marriage. Thus, the matching of split households with the original panel households is not a straightforward exercise. While doing any analysis the split households need to be handled carefully.
- (iii) Only the rural sampled households in Punjab and Sindh are covered in all three rounds, so the analysis of the three-wave data is restricted to these two provinces.
- (iv) For the analysis of all rural areas covering four provinces, panel data are available for the 2001 and 2010 rounds.

5. SCOPE OF THE PANEL SURVEY

The scope of the panel survey is examined in terms of the types of information (modules) gathered through the structured questionnaires. In all three rounds, two separate questionnaires for male and female respondents were prepared and different modules were included in these questionnaires (Table 3). A two-member team of

enumerators, one male and one female, visited each sampled household to gather information. Female enumerators were responsible to fill the household roster and pass it immediately to her male counterpart. Education and employment modules were included in both male and female questionnaires but the relevant information regarding children (under 5 years old), both male and female, was recorded in the female questionnaire. One major objective of the PRHS-PPHS panel survey has been to examine the movement into or out of poverty therefore a detailed consumption expenditure module has been a part of the female questionnaire in all the three rounds. Expenditures on durable items, however, were recorded in the male questionnaire. Health and migration modules were included in PRHS-I and PPHS 2010 rounds. A module on household-run businesses and enterprises was part of the latter two rounds as well.

Each round of the survey has had certain specific areas of focus. Agriculture, for example, was the main focus of the PRHS-I when information even at the plot level was collected from the land operating households. In the other two rounds only a brief agriculture module was included. The main focus of the PRHS-II was mental health, dowry, inheritance and marriage-related transfers. The PPHS-2010 was conducted at a time when inflation was high and the nation had also faced some natural disasters including droughts and floods. In the latest round modules on shocks, food security, subjective wellbeing and overall security were specially included in the questionnaire.

In short, the scope of the three rounds of the panel survey is wide. A variety of social, demographic and economic issues can be explored from these rounds. While some core modules are common to all rounds, there are others that are specific to a certain round. Some of the information is, thus, cross-sectional in nature but can be linked to the household socio-demographic dynamics made available through the core modules.

Table 3

Scope of the Panel Survey: Modules included in Household Questionnaires

Modules	PRHS-I (2001)		PRHS-II (2004)		PPHS (2010)	
	Male	Female	Male	Female	Male	Female
Household Roster	√	√	√	√	√	√
Education	√	√	√	√	√	√
Agriculture	√	×	√	×	√	×
Non-Farm Enterprises	√	×	×	×	√	×
Employment	√	√	√	√	√	√
Migration	√	×	√	×	√	×
Consumption	√	√	√	√	√	√
Credit	√	×	√	×	√	×
Livestock Ownership	×	√	×	√	×	√
Housing	×	√	×	×	×	√
Health	×	√	×	√	×	√
Dowry and Inheritance	×	√	×	√	×	×
Mental Health	×	×	×	√	×	×
Marital History and Marriage Related Transfers	×	×	×	√	×	×
Shocks and Coping Strategies	×	×	×	×	×	√
Household Assets	×	×	×	×	×	√
Household Food Security	×	×	×	×	×	√
Security	×	×	×	×	√	√
Subjective Welfare	×	×	×	×	√	√
Business and Enterprises	×	×	×	×	√	×
Transfer/Assistance from Programme and Individuals	×	×	×	×	√	×

6. AN ANALYSIS OF THE SAMPLE ATTRITION

As shown earlier, in the PRHS-PPHS data have been collected from the same households over three points of time- 2001, 2004 and 2010. It is common in such surveys that some participants (households) drop out from the original sample for a variety of reasons including geographical movement and refusal to continue being part of the panel. This attrition of the original sample represents a potential threat of bias if the attritors are systematically different from the non-attritors. It can lead to 'attrition bias' because the remaining sample becomes different from the original sample [Miller and Hollist (2007)]. If the participating units, however, are not dropped out systematically, meaning that there are no distinctive characteristics among the attriting units, then there is no attrition bias even though the sample has decreased between waves. It is, therefore, important to examine the attrition bias in our panel survey.

6.1. Theoretical Considerations¹

Attrition in panel surveys is one type of non-response. At a conceptual level, many of the insights regarding the non-response in cross-sections carry over to panels. According to Fitzgerald, *et al.* (1998), attrition bias is associated with models of selection bias. Their statistical framework for the analysis of attrition bias, which has been used by several other studies [see for example, Alderman, *et al.* (2000); Thomas, *et al.* (2001); Aughinbaugh (2004)], makes a distinction between selection of variables observed in the data and variables that are unobserved. Alderman, *et al.* (2000) believe that, 'if there is sample attrition, then it has to be seen whether or not there is selection of observables. Selection of observables includes selection based on endogenous observables, which occurs prior to attrition (e.g. in the first round of the survey). Even if there is selection of observables, this does not necessarily bias the estimates of interest. Thus, one needs to test for possible attrition bias in the estimates of interest as well' [Alderman, *et al.* (2000)].

Assume that the object of interest is a conditional population density $f(y/x)$ where y is scalar dependent variable and x is a scalar independent variable (for illustration, but in practice making x a vector is straightforward):

$$y = \beta_0 + \beta_1 x + \varepsilon, \text{ } y \text{ observed if } A=0 \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad (1)$$

where A is an attrition indicator equal to 1 if an observation is missing its value y because of attrition, and equal to zero if an observation is not missing its value y . Since (1) can be estimated only if $A=0$ that is, one can only determine $g(y|x, (A=0))$, one needs additional information or restrictions to infer $f(\cdot)$ from $g(\cdot)$, which can be derived from the probability of attrition, $PR(A=0|y, x, z)$, where z is an auxiliary variable (or vector) that is assumed to be observable for all units but not included in x . This leads us to the estimation of the following form:

$$A^* = \delta_0 + \delta_1 x + \delta_2 z + V \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad (2)$$

$$A = 1 \text{ if } A^* \geq 0 \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad (3) \\ = 0 \text{ if } A^* < 0$$

¹This sub-section depends heavily on Arif and Biquees (2006) who have examined the attrition bias between two rounds of the Pakistan Socio-Economic Survey (PSES) carried out in 1998-99 and 2001 by the Pakistan Institute of Development Economics.

If there is selection of observables, the critical variable is z , a variable that affects attrition propensities and is also related to the density of y conditional on x . In this sense, z is “endogenous to y ”. Indeed, a lagged value of y can play the role of z if it does not have structural relationship with attrition. Two sufficient conditions for the absence of attrition bias due to attrition of observables are either (1) z does not affect A or (2) z is independent of y conditional on x . Specification test can be carried out of either of these two conditions. One test is simply to determine whether candidates for z (for example, lagged value of y) significantly affect A . Another test is based on Beketti, *et al.* (1988), and is known as BGLW test. It has been applied by Fitzgerald, *et al.* (1998) and Alderman, *et al.* (2000). In the BGLW test, the value of y at the initial wave of the survey (y_0) is regressed on x and on A . This test is closely related to the test based on regressing A and x and y_0 (which is z in this case); in fact, two equations are simply inverses of one another [Fitzgerald, *et al.* (1998)]. Clearly, if there is no evidence of attrition bias from these specification tests, then one has the desired information on $f(y/x)$.

6.2. Extent of Attrition

Table 4 presents the attrition rate for different rounds. Between 2001 and 2010, the attrition rate was around 20 percent while the rate for the 2004 to 2010 period was 25 percent, suggesting some households had dropped in 2004 and re-entered the panel in 2010. For the 2004-10 period, the highest attrition rate is found in Balochistan hinting towards more movement of sampled households than in other provinces.

Table 4

Sample Attrition Rates of Panel Households—Rural

	2001-2004	2001-2010	2004-2010 (%)
Pakistan	14.1	19.6	24.9
Punjab	12.9	17.1	23.8
Sindh	15.7	18.3	26.2
KPK	—	16.1	—
Balochistan	—	33.2	—

Source: Authors’ computations based on PRHS 2001 and PPHS 2010 micro-datasets.

6.3. Attrition Bias

As stated earlier, the urban sample was included in the panel survey in 2010 for the first time and hence the attrition issue is related to the rural sample. It has also been noted that the PRHS-II was limited to two large provinces, Punjab and Sindh. All the rural areas were covered in round I (2001) and round III (2010). The attrition bias is examined between the two waves 2001 and 2010. Five models have been estimated where the dependent variable is whether attrition occurred between these two rounds (1 = yes; 0 = no), results for which are presented in Table 5. The sample used in these models consists of all 2001 households and all regressors are measured in 2001.

Table 5

Determinants of Attrition through Logit Regression

Correlates (2001/02)	Model 1	Model 2	Model 3	Model 4	Model 5
Log per capita consumption	-0.286*	-0.342*	-0.353*	-0.214**	-0.152***
Log household size		-0.257*	-0.177***	-0.014	0.056
Households with 1 or 2 family members only (yes=1)			0.416***	0.426***	0.353
Age of head of household (years)				0.001	0.003
Age-square of head of household				0.000	0.000
Female headed households (yes=1)				0.378	0.493***
Literacy of the head (literate=1)				-0.138	0.010
Livestock owned (yes=1)				-0.443*	-0.451*
land owned (yes=1)				-0.280*	-0.377*
	Provinces (Punjab as ref.)				
Sindh					-0.009
KPK					-0.021
Balochistan					0.910*
Constant	0.580	1.458**	1.36**	0.926	0.222
LR chi-square	11.93 (1)	19.35(2)	21.63(3)	53.71 (9)	102.63 (12)
Log likelihood	-1353.789	-1350.079	-1348.941	-1332.229	-1307.268
Observations	2,714	2,714	2,714	2,711	2,711

Source: Authors' computations based on PRHS 2001 and PPHS 2010 micro-datasets.

Note: ***P<0.01; ** P<0.05, * P<0.10.

Following Thomas, *et al.* (2001) and Arif and Bilquees (2006), the first model of attrition includes the only one covariate, $\ln(\text{PCE})$, where per capita consumption (PCE) is used as a measure of households' economic status. Table 5 presents coefficient estimates from the logit regressions. The first model indicates that there is a statistically significant negative relationship between PCE and the probability of leaving the panel. On average, lower economic status households were more likely to attrite between the two waves, so without weighting, the PPHS-2010 would be lesser representative of lower economic status households than would be a random household survey.

In model 2, two variables, $\ln(\text{PCE})$ and $\ln(\text{household size})$ have been included. Both PCE and family size (in 2001) are positively and significantly associated with a household staying part of the subsequent round of the panel survey. The third model in Table 5 adds one dummy, that of a household consisting of only one or two members. The association between attrition and PCE and household size still remains negatively significant. On the other hand, small size households (with 1 or 2 members) show a significant association with attrition.

Model 4 included measures related to three characteristics of the head of the household, which are age, sex and literacy. None of these variables turned out to be statistically significant. Two economic variables, ownership of livestock and land, and provincial dummies are added in model 5. Both the economic variables are significantly associated with keeping households part of the panel and maintaining them as non-attritors (see Table 5). Among the provinces, households in Balochistan are more likely to leave the sample than households located in other provinces. It is evident from the multivariate analyses that there is a positive association between leaving the panel and small household size. Improving economic status of the household is statistically significant to keep the household in the sample, so it is mainly the poorer households that are attriting.

As discussed in the beginning of this section, BGLW test, introduced and used initially by Beckett, *et al.* (1988), is the other method of testing the attrition bias. This test examines whether those who subsequently leave the sample are systematically different from those who stay in terms of their initial behavioural relationships. We estimate the consumption (lnPCE) equations as well as poverty equations, dividing the survey participants into two subsets—all 2001 households, and those still in the sample in 2010, labelled as ‘Always in’ or non-attritors.

Tables 6 and 7 present estimates of OLS regression for consumption equations and logit estimates for poverty equations respectively. A standard set of household and the head of the household characteristics, including age, and literacy of the head of the household, family size, and ownership of dwelling unit and livestock have been entered as independent variables into these equations. All the estimates are significant, as can be seen from Table 6 and Table 7. These estimates indicate a number of associations that are consistent with widely-held perceptions about consumption behaviour and poverty. For example, age and literacy of the head of the households have a positive impact on consumption while they are negatively associated with poverty. A similar pattern of association was also found for family size as it has a positive association with poverty but a negative relation with the per capita consumption expenditure. The ownership of both livestock and land has a positive association with per capita expenditure, but a negative relation with the incidence of poverty.

Table 6

Household Expenditure: OLS Regression Model 2001-2010

Variables	Full Sample		‘Always in’ (Non-attrition)		t-difference test
	Coefficients	St. Error	Coefficients	St. Error	
Age (years)	-0.001	0.004	0.001	0.004	-0.500
Age ²	0.000	0.000	0.000	0.000	0.000
Literacy (literate=1)	0.196*	0.023	0.190*	0.025	0.251
Family Size	-0.032*	0.003	-0.036*	0.003	1.333
Land Ownership (yes=1)	0.255*	0.023	0.252*	0.025	0.125
Livestock	0.142*	0.025	0.133*	0.028	0.341
Own House (yes=1)	-0.104**	0.047	-0.134**	0.055	0.592
Constant	6.838*	0.105	6.870*	0.117	-0.290
F-stat	56.46		47.66		-
R-square	0.1305		0.1367		-
Observations	2,642		2,115		-

Source: Authors’ computations based on PRHS 2001 and PPHS 2010 micro-datasets.

***P<0.01; ** P<0.05, * P<0.10.

Table 7

Correlates of Poverty: Logistic Regression Model 2001-2010

Correlates	Full Sample		'Always in' (Non-attritors)		t-difference test
	Coefficients	St. Error	Coefficients	St. Error	
Age (years)	0.025	0.019	0.022	0.022	0.147
Age ²	0.000***	0.000	0.000	0.000	0.000
Literacy (literate=1)	-0.545*	0.102	-0.504*	0.117	-0.376
Family Size	0.093*	0.011	0.108*	0.013	-1.257
Land Ownership (yes=1)	-0.827*	0.102	-0.840*	0.116	0.120
Livestock (yes=1)	-0.592*	0.105	-0.504*	0.122	-0.780
Own House (yes=1)	0.538**	0.210	0.639**	0.263	-0.430
Constant	-1.817*	0.483	-1.994*	0.568	0.339
LR chi-square	206.39		160.22		-
Log likelihood	-1374.198		-1058.706		-
Observations	2,642		2,115		-

Source: Authors' computations based on PRHS 2001 and PPHS 2010 micro-datasets.

*** P<0.01; **P<0.05; * P<0.1.

Our interest here, however, is more in the difference that the attritors might have made to the sample. To ascertain this we apply the t-difference test with the following hypotheses and assumption:

H₀: No significant difference between attritor and non-attritor.

H₁: Significant difference exists between attritor and non-attritor.

Assumption: unequal sample size, unequal variance.

The t-difference test results (see last columns of Table 6 and 7) show that there are no significant differences between the set of coefficients for the sub-sample of those missing in the follow-up versus the sub-sample of those re-interviewed for indicators of either consumption or poverty. These estimates, therefore, suggest that the coefficient estimates of standard background variables are not affected by sample attrition.

7. CONCLUSION

The PRHS-PPHS panel is a rich source of information regarding a range of socio-economic and demographic processes, and a means to understand their dynamics over time. Along with having a few core modules the panel questionnaire is flexible enough to accommodate any particular area of interest in a specific round without affecting the overall efficiency of the survey design. Addition of the urban sample in 2010 to the previously all rural sample has made the panel design even more comprehensive. With three rounds having been carried out so far, in 2001, 2004 and 2010, the panel sample retains its qualities despite all the attritions and the phenomenon of split households.

ANNEXURES

Table A1

Sample list for Pakistan Panel Household Survey 2010: Punjab

Province	Code	District	Code	Tehsil	Code	Village	Code				
Punjab	1	Faisalabad	1	Faisalabad	1	Saddon 206RB	1				
				Jaranawala	2	Sing Pura	2				
				Gojra	3	Jarwanwala Chak	3				
				Summandri	4	Subdarawala 363JB	4				
						Khalishabad 356JB	5				
						Summandri	6				
				Attock	2	Feth Jang	5	Khirala Kalan	5	7	
								Pindi Ghaip	6	Thathi Gogra	8
										Kareema	9
								Hattar	10		
								Makyal	11		
								Gulyal	13		
		Dhock Qazi	14								
		Hafizabad	5					Pindi Bhatian	11	Khatteshah	53
										Nasowal	54
										Khidde	55
				Bahoman	56						
				Daulu Kalan	57						
				Bagh Khona	58						
				Shah Behlol	59						
				Purniki	60						
				Thata Karam Dad	61						
				Mona	62						
		Vehari	6	Mailsi	12	Chak No 118-WB	63				
						Chak No 190 WB	64				
						Kot Soro	65				
						Chak No 195 WB	66				
						Mandan	67				
						Kot Muzzfar	68				
						Muradabad	69				
						Chak No 109 WB	70				
						Chak No 166- WB	71				
						Maqsooda	72				
		Punjab	1	Muzafar Garh	7	Ali Pur	13	Mail Manjeeth	73		
								Makhan Bela	74		
								Tibbah Barraha	75		
								Malik Arain	76		
								Kohar Faqiran	77		
								NauAbad	78		
								Kundi	79		
								Nabi Pur	81		
								Kotla Afghan	82		
Ghunia	83										
Bahawalpur	8			Ahmed Pur East	14	Chak No 157- N.P.	84				
						Haji Jhabali	85				
						Mad Rashid	87				
						Mukhawara	88				
						Pipli Rajan	89				
						Qadir Pur	90				
						Ladpan Wali	91				
						Chak Dawancha	92				

Table A2

Sample list for Pakistan Panel Household Survey 2010: Sindh

Province	Code	District	Code	Tehsil	Code	Village	Code												
Sindh	2	Badin	3	Badin	7	Kerandi	21												
						Golarchi	8	Kalhorki	22										
								Shaikhpur	23										
								Khoro	24										
								Khirdi	25										
								Bhameri	26										
				Walhar	27														
				Parharki		Golarchi	28												
						Lucky	29												
						Nurlut	30												
						Mitho Debo	31												
						Sorahdi	32												
						Chakri	33												
				Nawab Shah	9	Daulat Pur	15	Fatehpur		Mari Wasayo	34								
										Bajhshan		Kandiari	35						
												Khirion		Jagpal	36				
														Kandhari		Kandhari	37		
																Khar		Khar	39
		Sindal Kamal																Sindal Kamal	40
								Kaka										Kaka	93
										Bogri								Bogri	94
												Manhro						Manhro	95
														Uttar Sawri				Uttar Sawri	96
																Mir Pur Khas	10	Kot G. Mohammad	16
		Deh 320	98																
		Deh 346	99																
		Deh 339A	100																
		Deh 306	101																
		Deh 302	102																
		Larkana	11					Qamber Ali Rato Dero	17	Chacha		Deh 285	103						
												Rato Dero	18	Deh 257	104				
				Dera		Deh 306	105												
						Laktia								Deh 302	106				
														Do-Abo		Deh 285	107		
																Nather		Deh 257	108
										Haslla								Chacha	109
Sanjar Abro												Dera	112						
				Khan Wah								Laktia	113						
						Khuda Bux						Do-Abo	114						
												Naudero		Nather	115				
														Saidu Dero		Haslla	116		
			117																
	118																		
	119																		
	120																		
	121																		
	122																		

Table A3

Sample list for Pakistan Panel Household Survey 2010: Khyber Pakhtunkhwa

Province	Code	District	Code	Tehsil	Code	Village	Code			
KP	3	Dir	4	Blambut Adenzal	9	Katigram	41			
						Batam	42			
						Shah Alam Baba	43			
						Bakandi	44			
						Khanpur	45			
						Kamangara	46			
						Malakand	47			
						Khema	48			
						Khazana	49			
	Shehzadi	50								
	Munjal	51								
	Mardan	12	Takht Bhai	19	Khan Killi	125				
					Dagal	126				
					Jangirabad	127				
					Saidabad	129				
					Mian Killi	130				
					Fethabad	131				
					Seri Behial	133				
					L. Marwat	13	L. Marwat	20	Nar Akbar	135
									Nar Langar	136
	Alwal Khel	138								
Gorka	141									
						Ghazi Khel	142			

Table A4

Sample list for Pakistan Panel Household Survey 2010: Balochistan

Province	Code	District	Code	Tehsil	Code	Village	Code
Balochistan	4	Loralai	14	Loralai	21	Sanghri	145
						Urd Shahboza	146
						Sor Ghand	147
						Nigang	148
						Marah Khurd	149
		Khuzdar	15	Khuzdar	22	Mekhtar	150
						Tor	151
						Bajori Kalan	153
						Ghorawah	154
						Bhat	155
		Gawadar	16	Gawadar	23	Khat Kapper	156
						Sabzal Khan	157
						Khorri	159
						Par Pakdari	160
						Ankra	161
						Chibab Rekhani	162
						Dhorgati	163
						Grandani	164
						Nigar Sharif	165
						Shinkani Dar	167
Sur Bandar	168						

REFERENCES

- Afzal, M. and T. Ahmed (1974) Limitations of Vital Registration System in Pakistan against Sample Population Estimation Project: A Case Study of Rawalpindi. *The Pakistan Development Review* 13:3.
- Alderman, H., J. Behrman, H. Kholer, J. Mauccio and S. Watkins (2000) Attrition in Longitudinal Household Survey Data: Some Tests for Three Developing Country Samples. The World Bank, Development Research Group Rural Development. (Policy Research Working Paper 2447).
- Arif, G. M. and F. Bilquees (2006) An Analysis of Sample Attrition in PSES Panel Data. Pakistan Institute of Development Economics, Islamabad. (MIMAP Technical Papers Series No. 20).
- Aughinbaugh, A. (2004) The Impact of Attrition on the Children of the NLSY97. *The Journal of Human Resources* 39:2.
- Beckett, S., W. Gould, L. Lillard, and F. Welch (1988) The Panel Study of Income Dynamics after Fourteen Years: An Evaluation. *Journal of Labour Economics* 6.
- Fitzgerald, J., P. Gottschalk, and R. Moffit (1998) An Analysis of Sample Attrition in Panel Data. *The Journal of Human Resources* 33:2.
- Miller, R. and C. Hollist (2007) *Attrition Bias*. Department of Child, Youth and Family Studies, University of Nebraska-Lincoln.
- Thomas, D., E. Frankenberg, and J. Smith (2001) Lost but not Forgotten: Attrition in the Indonesian Family Cycle Survey. *The Journal of Human Resources* 36:3, 556–592.