

Pakistan Panel Household Survey (PPHS)

1. Introduction

The primary goal of the Pakistan Panel Household Survey (PPHS), a joint project of Pakistan Institute of Development Economics (PIDE) and the World Bank, is to create a longitudinal and multi-thematic database to analyze poverty and social dynamics in Pakistan. This note first outlines the main features of the Pakistan Panel Household Survey (PPHS), and then it provides a guidance for panel data users how to merge and match the micro-data files cross-sectional or across the three rounds of the survey. Merging of data files can be done at the household level and/or individual level.

2 Geographic coverage of the PPHS

International Food Policy Research Institute (IFPRI) conducted a panel household survey in rural areas of four districts over a period of five years from 1986 to 1991 (Dir in Khyber Pakhtunkhwa; Attock and Faisalabad in Punjab; and Badin in Sindh). In 2001, the joint project of the PIDE and World Bank took a major initiative not only to revisit the IFPRI panel households in four districts after a gap of 10 years, but also expanded the sample from four to 16 districts by adding 12 new districts from four provinces of the country (Table 1). Since the sample was limited to rural areas of the 16 districts, it was named as the Pakistan Rural Household Survey (PRHS-I) 2001.

The second round of the PRHS (II) was carried out in 2004, but only in two provinces Punjab and Sindh; Khyber Pakhtunkhwa (KP) and Balochistan could not be covered due to security reasons. The third round was completed in 2010 covered all 16 districts (Map 1). While the first two rounds were rural only, an urban sample was added in the third round and therefore renamed as the Pakistan Panel Household Survey (PPHS)¹.

¹ For sample detail see Nayab, Durr-e- and G. M. Arif (2014). Pakistan Panel Household Survey Sample Size, Attrition and Socio-demographic Dynamics (data note), *Pakistan Development Review (special issue)*, vol. 53(2), 2014.

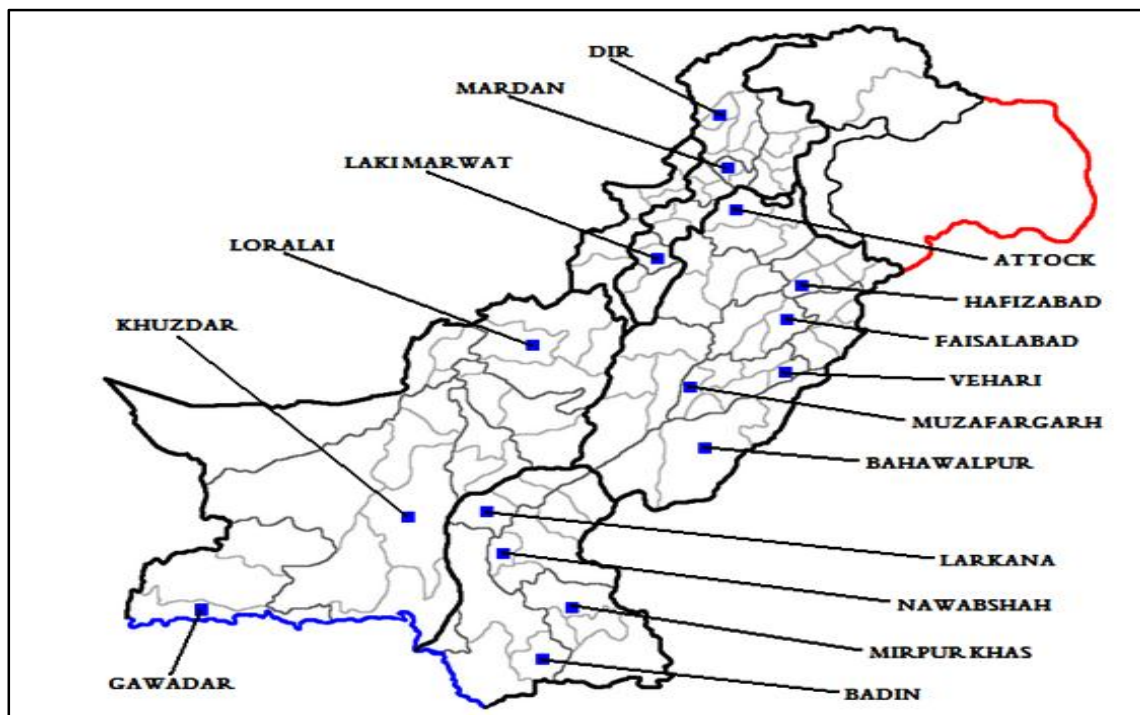
Table 1: Primary Sampling Units (PSUs) by Province and District

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
	Mir PurKhas ^b	8	4
	Larkana ^b	11	7
NWFP	Dir ^a	11	2
	Mardan ^b	7	6
	LakiMarwat ^b	5	2
Balochistan	Loralai ^b	7	2
	Khuzdar ^b	7	3
	Gawadar ^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

Figure 1: Map Showing Selected Districts for the PPHS-2010



3 Sample attrition

Attrition in panel surveys is one type of non-response. Attrition bias can affect the findings of the subsequent rounds of a panel survey. Sample attrition in the PPHS has been analyzed by Nayab and Arif (2014).² Based on various socio-demographic and economic indicators, they found the attrition random, suggesting that panel sample retains its qualities despite all the attritions and inclusions of split households.

4 Handling the Split Households

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 there 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 not handled properly the demographic composition of the sampled households is likely to change over time.

The PRHS-II carried out in 2004 did not track the households split from the first round sample (PRHS-I) conducted in 2001. However, 292 sister-in-law households were interviewed. In the third round (PPHS) carried out in 2010, split households were interviewed. The interviewed split households, however, were the ones that were residing in the same village as the original panel household. In other words, movement of panel households or its members out of the sampled villages were not followed because of the high costs involved in this type of follow-up.

In short, four features of the three rounds of the panel data are noteworthy, which are as follows:

- i. Urban households, which have been included first time in the sample in the third round (PPHS) held in 2010, are not panel households as yet. 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

² Nayab, Durr-e- and G. M. Arif (2014), Pakistan Panel Household Survey Sample Size, Attrition and Socio-demographic Dynamics (data note), *Pakistan Development Review (special issue)*, vol. 53(2), 2014.

households is not a straight forward exercise. While doing any analysis the split households need to be handled carefully.

- iii. Only rural sampled households in Punjab and Sindh are covered in all three rounds so an analysis of the three-wave micro-data will be restricted to these two provinces.
- iv. For the analysis of all rural areas covering four provinces, panel data are available for two rounds only: PRHS-I (2001) and PPHS (2010).

5 Survey questionnaires and Scope of the Panel Survey

In all three rounds, two separate questionnaires for male and female respondents were prepared and different modules were included in these questionnaires. A two-member team of enumerators, one male and one female, visited each sampled household to gather information through these questionnaires. 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 both male and female children (under 5 years old) were recorded in the female questionnaire.³

Table 2 reports the modules included in three rounds of the survey. One major objective of the panel survey has been to examine the movement into or out of poverty therefore a detail 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 (2001) 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 (2001) when information at even the plot level was collected from land operating households. In the other two rounds only a brief agriculture module was included. The main focus of the PRHS-II (2004) was mental health, dowry, inheritance and marriage related transfers.

³ All questionnaires used in three rounds of the panel survey can be downloaded from the PPHS-PIDE website.

Table 2: Modules covered in Three Rounds

Modules	PRHS-1 (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 program and individuals	×	×	×	×	√	×

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, and subjective wellbeing 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.

6 Merging Micro-data: Cross-Sectional and/or across three rounds?

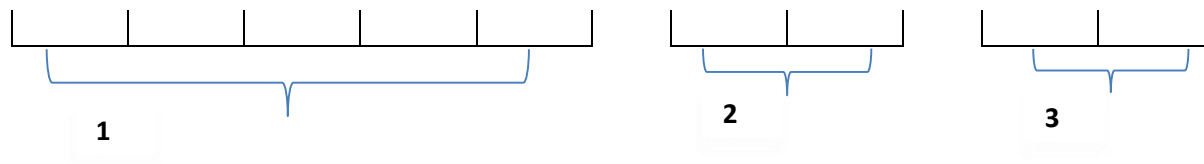
Multiple modules have been covered in three rounds of the panel survey. Micro-data is stored in separate male and female folders, as extracted from male and female questionnaires. Each male and female folder comprises of a number of files. Like other longitudinal surveys, it is very important to identify the households and individuals that were interviewed in various rounds of the panel survey. Both the households and individuals covered in different rounds may be cross-sectional or panel. In order to facilitate the process of merging micro-data from various rounds at household level and/or individual level, this section gives information first on “**Household Identifier**” and then on “**Individual Identifier**”.

Household Identifier

In order to merge different cross-sectional files, a numeric variable “compno” has been used in PRHS-I (2001) and PRHS-II (2004) rounds for household identification. The “compno” variable exists in all the files of these two rounds. The “compno” variable range from 1 to 4 digits. The last or third round, PPHS (2010), has a 10-digit variable “idcode” in all micro-data files, showing the household identification for cross-sectional analysis.

Identifier	PRHS-1 (2001)	PRHS-II (2004)	PPHS (2010)
Cross-sectional Household Identifier	compno	compno	idcode
Panel Household Identifier	panel_compno	panel_compno	panel_compno

For panel merging of various rounds, a variable “panel_compno” has been added in the cover pages and roster modules of all three rounds for panel household identification. “panel_compno” comprises of 9-digits with the following 3 specifications:



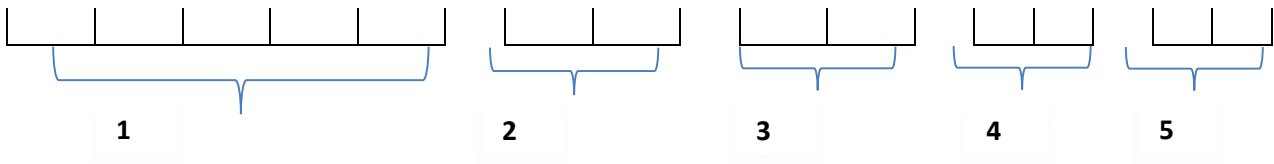
1. The first five digits are the numbers that identify the panel household or the computer number of a household.
2. The 6th and 7th digit indicates the round when the household was first time interviewed/formed. “01” mean that the household was first time interviewed/formed in first round (2001), “02” mean that the household was first time interviewed/formed in 2nd round (2004) and “03” mean that the household was first time interviewed/formed in 3rd round (2010).
3. The last two digits allow to identify the new households that are formed due to split-offs of panel members from their original household of 2001 (PRHS-I) and 2004 (PRHS-II). In the case of panel households, these two digits are “00”. Even in the case of new households other than split-offs, this number appears to be “00” i.e. the urban sample was included in 3rd round (PPHS-2010) of the panel survey. In case of new split-offs households, these two digits correspond to the individual identifier of the panel member who split from the original household, “01” appears to be first split-off, “02” for 2nd and so on. It is worth noting that in the third round (PPHS-2010), there are up to 6 split-off households – split or separated from an original panel household of 2001 (PRHS-I) or 2004 (PRHS-II).

Individual Identifier

In order to merge different cross-sectional data files at individual level, a numeric variable “pide01” has been used for 2001 round or PRHS-I, “pide04” for 2004 round or PRHS-II, and “pide10” for 2010 round or PPHS. These individual identifiers exists in all files of these three rounds 2001, 2004 and 2010 in which individual information have been gathered. All the three individual identifiers comprise of 1 to 2-digit numeric numbers.

Identifier	PRHS-1 (2001)	PRHS-II (2004)	PPHS (2010)
Cross-sectional Individual Identifier	pid01	pid04	pid10
Panel Individual Identifier	panel_pide	panel_pide	panel_pide

For panel merging of various rounds of PPHS at individual level, we have added a variable “panel_pide” in all the roster modules of the three rounds for panel household identification; this “panel_pide” can be used to trace out individuals over time. The **panel_pide** is a unique individual indicator that stays constant throughout rounds. The **panel_pide** is comprised of 13 digits. The first 9 digits correspond to the household number which has been mentioned above.



1. The 10th and 11th digit indicates the round when the individual was first time interviewed/join house e.g. “01” means that the individual was the part of house in first round (2001), “02” mean that the he/she joined in the 2nd round (2004) and “03” means that he/she joined in the 3rd round (2010).
2. The 12th and 13th digits indicate the individual *idcode* (pid01, pid04 and pid10) when individual was first time interviewed.

The example below explains how the **Household Identifier** and **Individual Identifier** can be used while working on micro-data files cross-sectional or across three rounds..

Panel Household in 2001

Suppose a households was first time interviewed in 2001, having four members where Ashraf and Tahira are husband and wife. They have 2 children, Rabia and Ali. The cross-sectional household key “compno” and individual key “pid01” will be used for cross-sectional merging. The table below shows the 9-digit panel household key “panel_compno”, where 6th and 7th digit indicates

the round when the household was first time interviewed, while 8th and 9th digit indicates whether the household is panel or split. In table below, all these 9-digits of panel_compno are the same in panel_pide; its 10th and 11th digit indicates the round when the individual was first time interviewed, while 12th and 13th digit show the individual *idcode* (pid01) when individual was first time interviewed.

Compno	Name	pid01	panel_compno	panel_pide
1	Ashraf	1	10001 01 00	10001 01 00 01 01
1	Tahira	2	10001 01 00	10001 01 00 01 02
1	Rabia	3	10001 01 00	10001 01 00 01 03
1	Ali	4	10001 01 00	10001 01 00 01 04

Addition of Member in 2004 of Panel Household

Suppose another baby “Ahmed” born in the above discussed household after 2001 and before 2004 round. Moreover, some individuals’ cross-sectional key was also changed, i.e. Rabia has “3” in 2001 but “4” in 2004, Ali has “4” in 2001 but “3” in 2004. Table below shows that despite of changing cross-sectional key in 2004, the “panel_pide” of both Rabia and Ashraf is same for both 2001 and 2004 rounds. In fact 2001 “panel_pide” has been placed in front of those individuals which exist in 2001 and were interviewed in 2004. Ahmed was born between the two rounds, therefore, his 10th and 11th digit indicates that he has been interviewed in 2nd round (02) and the last two-digit indicate his pid04. The cross-sectional household key “compno” and individual key “pid04” will be used for cross-sectional merging.

compno	Name	pid04	panel_compno	panel_pide
1	Ashraf	1	10001 01 00	10001 01 00 01 01
1	Tahira	2	10001 01 00	10001 01 00 01 02
1	Ali	3	10001 01 00	10001 01 00 01 04
1	Rabia	4	10001 01 00	10001 01 00 01 03
1	Ahmed	5	10001 01 00	10001 01 00 02 05

Changes in the Panel Households from 2004 to 2010 and Split of Households

Suppose Ali, who was member of a panel household in 2004, married between 2004 and 2010 and started living in a separate household with his wife Adeela and a newborn child Saqib. This formation of a new household is called split household. The below tables show the establishment of key for both the original and split households - how the split-household key at household and

individual level is formed in 2010. The cross-sectional household key “idcode” and individual key “pid10” will be used for cross-sectional merging for both the panel and split households. It is worth noting that cross-sectional individual key of Ali has been changed overtime (see pid01, pid04 and pid10 of Ali). The household panel key “panel_compno” shows that only the last two digits have been changed, now “01” appears by showing that it is the first split of the panel household who was first time interviewed in 2001. Despite of formation of new house, the individual panel key of Ali is the same of 2001 and 2004. Adeela and Saqib were first time interviewed in 3rd round (2010), therefore they have “03” at 10th and 11th digit, while their last two digit indicate their pid10.

Original household

idcode	Name	Pid10	panel_compno	panel_pide
1010150104	Ashraf	1	10001 01 00	10001 01 00 01 01
1010150104	Tahira	2	10001 01 00	10001 01 00 01 02
1010150104	Rabia	3	10001 01 00	10001 01 00 01 03
1010150104	Ahmed	4	10001 01 00	10001 01 00 02 05

Split household

idcode	Name	pid10	panel_compno	panel_pide
1010150110	Ali	1	10001 01 01	10001 01 00 01 04
1010150110	Adeela	2	10001 01 01	10001 01 01 03 02
1010150110	Saqib	3	10001 01 01	10001 01 01 03 03