Appendix I—IHDS: The Design

One of the most important goals of this research is to deepen our understanding of human development in India. Unlike the large body of empirical literature that relies on aggregated secondary data for analysis of human development issues, in this study we have used data from the IHDS 2005, administered to a nationally representative sample of households. This appendix describes our data collection and sample selection methods, assesses the quality of data, and provides an overview of the data analysis techniques used in the preceding chapters. The authors of this monograph are designers and organizers of this survey. Data collection for this survey was supported by two grants (RØ1HDØ41455 and RØ1HDØ46166) from the US National Institutes of Child Health and Human Development with supplementary funding from the World Bank.

A survey that encompasses a full range of human development issues faces practical challenges, not encountered by more limited focus projects. Every issue, from questionnaire design, to data cleaning, to statistical analysis, is complicated by the decision to broaden the range of the human development issues addressed. The analytic gains are substantial, but the practical costs are also real. After a careful consideration of these issues it was decided to field the IHDS to over 41,000 households residing in rural and urban areas, selected from 33 states and union territories. The sample extends to 384 out of 593 districts identified in the 2001 Census. While financial and management limitations precluded inclusion of all districts in the sample, the selection of 384 out of 593 districts allows for a highly diverse sample. All states and union territories are included in the sample, with the exception of Andaman and Nicobar Islands, and Lakshadweep. These two contain less than 0.05 per cent of India's population and their island location, as well as requirement of special permits to visit some parts, make them difficult to survey.

The IHDS benefited from a rich history of survey research in India, generally, and from NCAER and its collaborating institutions, in particular. The questionnaire design was borrowed, as needed, from Indian and international household surveys. Some of the important Indian sources include the NSSs, the *NFHSs*, and the 1994 Human Development Profile of India. International sources include five countries and the Status of Women and Fertility Survey, the World Bank Living Standard Measurement Surveys, and Indonesian and Malaysian Family Life Surveys. Organization of fieldwork and oversight was in the capable hands of professionals with a generation of practical experience, culled from a wide variety of surveys. Data cleaning and analysis enlisted a small army of personnel with well developed, often obsessive, attention to detail. At its best, most of this work is invisible, thus, permitting the analyst and the reader to focus on the central research questions. But the success of those analyses and the validity of their conclusions depend on the competent execution of the survey itself. This chapter reviews the major issues of that execution.

SAMPLING

The IHDS is a nationally representative survey of 41,554 urban and rural households. It covers all states and union

territories of India, with the exception of Andaman, Nicobar, and Lakshadweep islands. These households are spread across 33 states and union territories, 384 districts, 1,503 villages and 971 urban blocks, located in 276 towns and cities. Districtwise coverage for total, rural, and urban sample is shown in Figures AI.1, AI.2 and AI.3.

These 41,554 households include 215,754 individuals. Statewise distribution of sampled households and individuals is presented in Table AI.1.

Villages and urban blocks (comprising of 150–200 households) formed the primary sampling unit (PSU) from which the households were selected. Urban and rural PSUs were selected using a different design. In order to draw a random sample of urban households, all urban areas in a state were listed in the order of their size with number of blocks drawn from each urban area allocated based on probability proportional to size. Once the numbers of blocks for each urban area were determined, the enumeration blocks were selected randomly with help from the Registrar General of India. From these Census Enumeration Blocks of about 150–200 households, a complete household listing was conducted and household samples of 15 households per block were selected.

The rural sample contains about half the households that were interviewed initially by NCAER in 1993–4 in a survey titled Human Development Profile of India (HDPI),[1] and the other half of the samples were drawn from both districts surveyed in HDPI as well as from the districts located in the states and union territories not covered in HDPI. The original HDPI was a stratified random sample of 33,230 households,

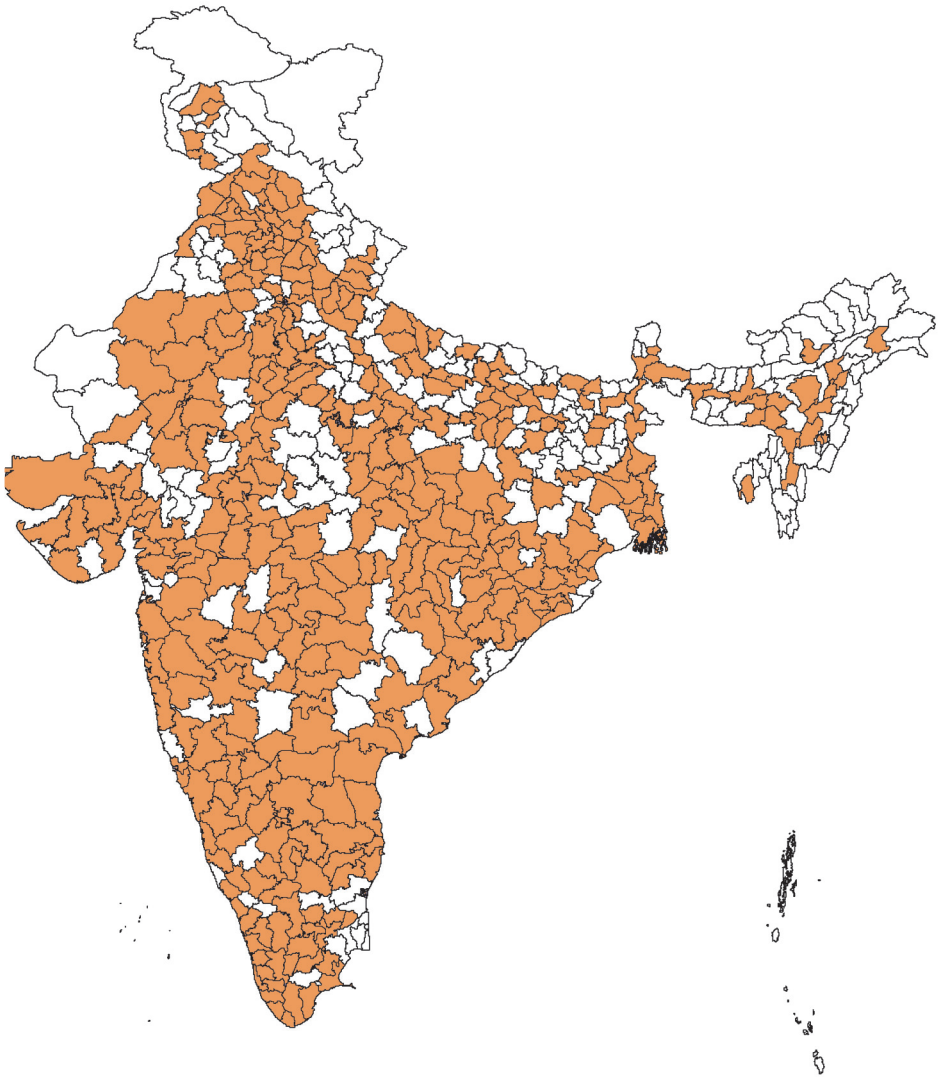

Figure AI.1 India Human Development Survey 2005, District Coverage—Urban and Rural Sample

Source: IHDS 2004–5.

[1] Shariff (1999).

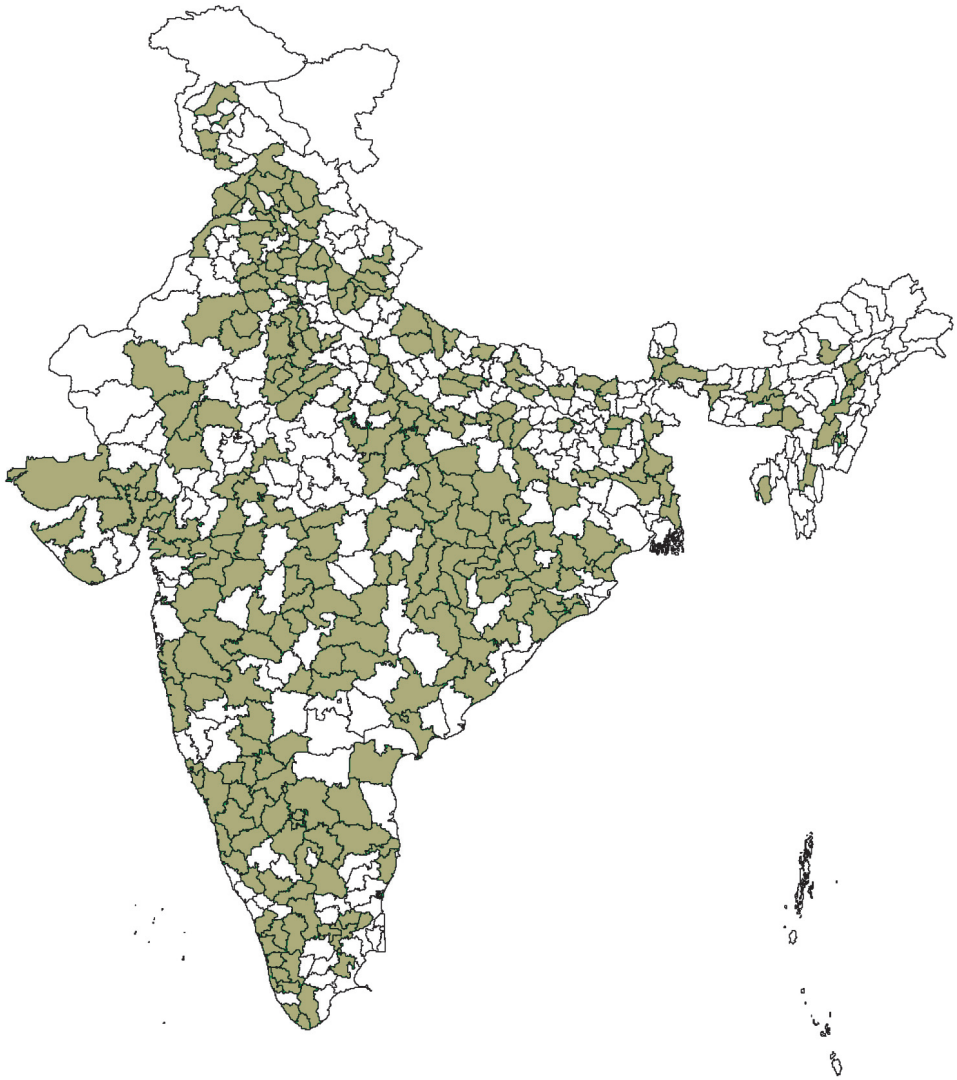

Figure AI.2 India Human Development Survey 2005, District Coverage—Rural Sample

Source: IHDS 2004–5.

located in 16 major states, 195 districts, and 1,765 villages. In states where the 1993–4 survey was conducted and re-contact details were available, 13,593 households were randomly selected for re-interview in 2005.

After a gap of 11–12 years, about 82 per cent of the households were contactable for re-interview, resulting in a resurvey of 11,153 original households, as well as 2,440 households which were separated from these root households, but were still living in the village. Distribution of the sample is described in Figure AI.4.

In order to check the representativeness of the sample, in each district, where re-interviews were conducted, two fresh villages were randomly selected using the probability proportional to size technique. In the villages selected for survey in this manner, 20 randomly selected households were interviewed. Comparing the panel sample with this randomly selected refresher sample, allows us to determine whether this panel sample is overrepresented among certain segments of the society. Table AI.2 compares the characteristics of the re-interview sample with the refresher sample for the districts where any re-interviews took place.

The comparison suggests that on most variables of interest such as caste, religion, education, and economic status, the re-interviewed sample does not differ substantially from the fresh sample.

Additionally 3,993 rural households were randomly selected from the states where the 1993–4 survey was not conducted, or where re-contact information was not available. This approach to combining a randomly selected panel sample, while refreshing it, with another random sample has been used in a variety of surveys including the Panel Study of Income Dynamics in the US and Malaysian Family Life

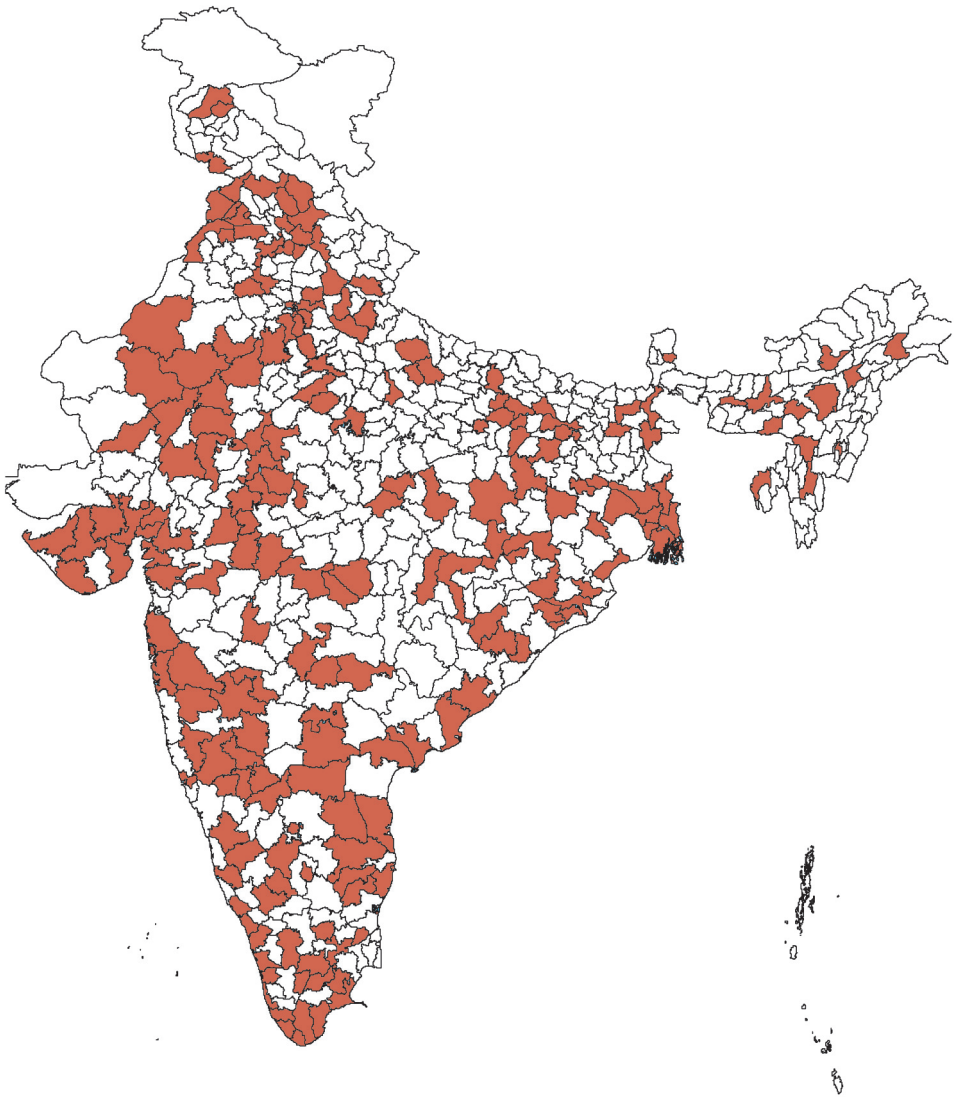

Figure AI.3 India Human Development Survey 2005, District Coverage—Urban Sample

Source: IHDS 2004–5.

Survey.[2] However, given the potential for significant sampling and non-sampling errors, we present a detailed analysis of the quality of IHDS data below.

COMPARATIVE RESULTS

IHDS was not intended to provide national nor, certainly, statewise estimates of levels of human development outcomes. There are already many excellent Indian surveys that fill that mission. The main purpose of IHDS is to provide a means for gaining insight by analysing the relationships among these human development outcomes and the connections between human development and its background causes.

Nevertheless, it is useful to compare IHDS estimates of human development levels with estimates from other more narrowly focused surveys that usually have larger sample sizes and smaller sampling errors. The NSSs and the *NFHSs* are obvious comparisons because of their excellent quality and wide use. The Indian Census provides another useful reference. The Census and these surveys differ not only in their objectives and design, but their question wording, sampling design, coding decisions, and government sponsorship, all of which should be expected to provoke somewhat different answers from respondents, and yield different frequencies (Table AI.3).

[2] Leslie Kish and Alastair Scott were the first to describe the probability sampling procedures which are designed to optimize the reselection or retention of sample units during a transition from an old to a new sample design. A description of this can be found in 'Retaining units after changing strata and probabilities', in the *Journal of the American Statistical Association*, Vol. 667, Number 335, Applications Section, September 1971.

Table AI.1 Statewise Distribution of IHDS Sample

	Disctrict in 2001 Census	Included in IHDS				Households Surveyed			Individuals Surveyed		
		Districts	Urban Areas	Blocks	Villages	Rural	Urban	Total	Rural	Urban	Total
Jammu and Kashmir	14	5	5	21	20	400	315	715	2,528	1,702	4,230
Himachal Pradesh	12	9	7	21	52	1,057	315	1,372	5,663	1,503	7,166
Punjab	17	13	11	36	61	1,033	560	1,593	6,202	2,831	9,033
Chandigarh	1	1	1	6	0	0	90	90	0	383	383
Uttaranchal	13	6	3	9	20	309	149	458	1,757	736	2,493
Haryana	19	14	6	18	79	1,350	268	1,618	8,112	1,291	9,403
Delhi	9	10	7	56	6	60	900	960	329	4,291	4,620
Rajasthan	32	23	17	60	88	1,590	895	2,485	9,663	4,805	14,468
Uttar Pradesh	70	43	24	75	138	2,389	1,123	3,512	14,966	6,499	21,465
Bihar	37	17	10	31	61	965	465	1,430	5,950	2,856	8,806
Sikkim	4	1	1	3	3	60	45	105	293	212	505
Arunachal Pradesh	13	1	1	3	6	120	45	165	623	209	832
Nagaland	8	4	1	2	5	100	30	130	480	84	564
Manipur	9	3	1	3	3	60	45	105	359	239	598
Mizoram	8	1	1	3	3	60	45	105	263	239	502
Tripura	4	2	1	3	7	184	45	229	818	190	1,008
Meghalaya	7	3	1	3	6	116	45	161	505	250	755
Assam	23	8	7	21	38	699	318	1,017	3,286	1,404	4,690
West Bengal	18	14	21	75	66	1,247	1,133	2,380	6,170	4,788	10,958
Jharkhand	18	6	9	27	26	519	405	924	2,913	2,095	5,008
Orissa	30	26	13	40	84	1,464	600	2,064	7,710	2,886	10,596
Chhattisgarh	16	15	6	18	49	905	270	1,175	4,833	1,377	6,210
Madhya Pradesh	45	31	13	42	121	2,177	628	2,805	12,392	3,409	15,801
Gujarat	25	17	14	60	70	1,167	911	2,078	5,926	4,234	10,160
Diu and Daman	2	2	0	0	3	60	0	60	281	0	281
Dadra and Nagar Haveli	1	1	0	0	3	60	0	60	315	0	315
Maharashtra	35	27	18	75	115	2,078	1,125	3,203	10,881	5,721	16,602
Andhra Pradesh	23	19	18	60	94	1,526	909	2,435	6,669	3,992	10,661
Karnataka	27	26	21	78	144	2,832	1,189	4,021	14,184	5,675	19,859
Goa	2	2	1	3	6	100	65	165	475	307	782
Lakshadweep	1	0	0	0	0	0	0	0	0	0	0
Kerala	14	12	14	42	61	1,089	642	1,731	4,892	3,089	7,981
Tamil Nadu	30	21	22	74	62	898	1,200	2,098	3,691	4,855	8,546
Pondicherry	4	1	1	3	3	60	45	105	245	228	473
Andaman and Nicobar	2	0	0	0	0	0	0	0	0	0	0
Total	593	384	276	971	1503	26,734	14,820	41,554	1,43,374	72,380	2,15,754

Source: IHDS 2004–5 data.

Table AI.2 Comparison of New and Re-interview Rural Sample in Districts Where Any Re-interviews Took Place

	New Sample	*Re-interview Sample*
Individual Characteristics		
Age		
0–4	10	9
5–9	12	11
10–14	12	13
15–19	10	10
20–9	17	16
30–9	14	13
40–59	17	18
60+	8	9
Sex		
Male	51	51
Female	49	49
Education		
Illiterate	44	44
1–4 Std	17	17
5–9 Std	27	27
10–11 Std	6	7
12 Some college	3	3
College graduate	2	2
Household Characteristics		
Social group		
Forward Caste Hindu	16	18
OBC	38	35
Dalit	23	26
Adivasi	12	10
Muslim	9	9
Christian, Sikh, Jain	2	2
Place of Residence		
Metro	0	0
Other urban	1	1
More developed village	50	45
Less developed village	49	54
Maximum Adult Education		
Iliterate	30	29
1–4 Std	10	10

(*contd*)

(Table AI.2 *contd*)

	New Sample	*Re-interview Sample*
5–9 Std	34	33
10–11 Std	11	12
12 Some college	8	8
College graduate	7	8
Household Income		
Negative—Rs 999	3	3
1st Quintile (Rs 1,000–14,000)	27	23
2nd Qunitile (Rs 14,001–22,950)	24	23
3rd Quintile (Rs 22,951–36,097)	19	21
4th Qunitile (Rs 36,098–69,000)	17	18
5th Qunitle (Rs 69,001+)	10	12

Source: IHDS 2004–5 data.

However, a comparison of IHDS data with the NSS (2004–5), *NFHS-III* (2005–6) and Census (2001) presented in Table AI.3 provides considerable reassurance about the robustness of IHDS data. IHDS sample distribution on urban residence, caste, and religion is remarkably similar to NSS and *NFHS-III*, although all three surveys (IHDS, NSS, and *NFHS*) have a higher proportion of households claiming Scheduled Caste status than enumerated in Census. The IHDS has a slightly higher proportion of households falling in Scheduled Caste category and slightly lower proportion in Scheduled Tribe category than NSS or *NFHS*. On other variables of interest, we find literacy and school enrolment in IHDS to be very similar to that in NSS. On work participation rate for males, IHDS falls in between NSS and Census estimates. However, given the special effort made to obtain estimates of women's unpaid work, it is not surprising that IHDS estimates for women's work participation are higher than both NSS and the Census. Family size estimates range from 4.7 in NSS to 5.3 in the Census. The average family size in IHDS was 5.2. Of particular interest is the poverty rate estimated at 25.7 per cent by IHDS, close to 27.5 per cent estimated by NSS. The IHDS records a higher proportion of households owning TV, using electricity and LPG gas than the NSS, possibly due to differences in question wording. But on most other variables, the IHDS results seem to be fairly consistent with the results from other surveys.

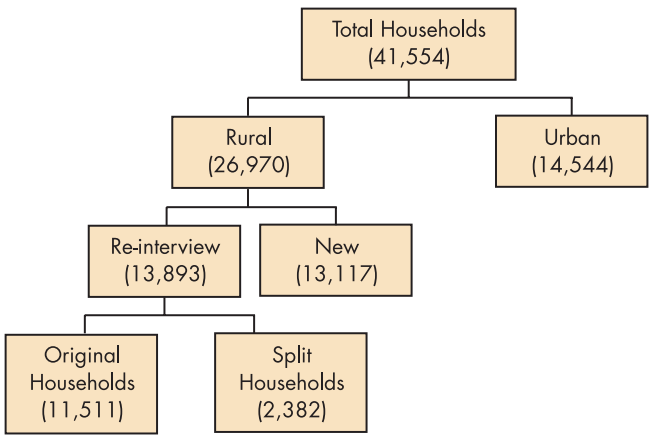

Figure AI.4 Sample Distribution

Note: 276 households were selected as rural but became urban by 2001, bringing the total of urban households to 14,820.

Source: IHDS 2004–5 data.

However, it is important to note that these broad similarities between IHDS data and other data sources do not remain quite so robust when we look at sub-national levels. Hence, we caution the readers about over interpreting IHDS estimates for statewise or other smaller samples. The IHDS sample sizes are large enough to investigate the general patterns that determine human development outcomes, but if readers desire a precise point estimate of the level of some particular indicator for a sub-sample of the Indian population, they are better referred to sources such as the NSS or the Census.

QUESTIONNAIRE DESIGN

The 100 pages of questions used in IHDS were carefully selected from items successfully administered in previous surveys in India and other developing countries, although some were modified after fielding these in the pre-testing of IHDS questionnaire. Some topics on which IHDS has special perspective (for example, marriage and gender relations) required the development of a new set of questions. But all questions, even those adopted from previous work, went through rigorous pre-testing and screening. The final

Table AI.3 Comparison of IHDS Estimates with Other Data Sources

	IHDS 2004–5	*NFHS–III 2005–6*	*NSS 2004–5*	*Census 2001*
Urban	26	31	25	28
Per cent literate				
Age 5+	67	67	66	NA
Age 7+	68	69	67	65
Caste				
Other Backward Classes	42	40	41	NA
Scheduled Castes	21	19	20	16
Scheduled Tribes	7	8	9	8
Other	30	32	31	NA
Religion				
Hindu	80	82	82	81
Muslim	14	13	13	13
Christian	2	3	2	2
Sikh	2	2	2	2
Buddhist	1	1	1	1
Jain	1	1	1	1
Others	2	1	1	1
Per cent currently in school (age 5–14)	80	NA	83	NA
Knowledge of AIDS (women)	54	61	NA	NA
Work participation rate for males	53	NA	55	52
Work participation rate for females	32	NA	29	26
Average family size	5	5	5	5
Number of children ever born to women (age 40–4)	4	4	NA	NA
Number of children ever born to women (age 45–9)	4	4	NA	NA
Per cent women married (age 15–49)	73	75	76	77
Per cent women married (all ages)	48	47	48	48
Per cent electricity	72	68	65	56
Per cent piped water	40	25	41	37
TV ownership (colour or b/w)	48	(Colour) 25	37	24
LPG use	33	25	22	18
Per cent flush toilets	23	NA	19	18
Per cent poor	26	NA	27	NA

Note: NA—not available due to potential measurement errors and/or small sample sizes.
Source: IHDS 2004–5 data.

questionnaires were the result of a careful, often painful, process of selection and revision in order to keep the questions understandable by respondents as well as the interview length manageable, with an eye on minimizing their burden as far as possible, without sacrificing the required detail.

Some parts of the questionnaire attempted to replicate other works as precisely as possible in order to maximize comparability. The consumption questions used for calculation of poverty incidence in Chapter 3, for instance, were copied from the short form of the consumption module

developed for NSS employment/unemployment survey. The 61st Round NSS survey was administered in 2004–5, allowing us to test the reliability of the IHDS estimates. One goal of IHDS was to compare a household's relative position on this much used consumption index with data on income and on household amenities, two other measures of economic position. Comparability required replicating the NSS measures as far as possible.

Other parts of the interview borrowed substantially from past work, but had to be adapted to the IHDS format. Reading, writing, and arithmetic tests were developed in conjunction with PRATHAM, although adapted for IHDS use. Since PRATHAM's *Annual Status of Education Report* was prepared in 2005, once again, this allows for considerable data quality evaluation and comparability.

Some often used questions had to be replaced by alternatives that our respondents found more understandable. The social network questions used in analysis, reported in Chapter 13, for example, first used a relational format—with whom do you talk when you seek advice—but were changed to a more direct, although less common positional format—do you know anyone in … which our respondents found easier to answer (and proved to be far more easily coded for analysis).

Some questions, even those used in many previous surveys, proved too ambiguous in pre-testing and had to be deleted altogether because no suitable alternative could be devised. For example, a question on interpersonal trust, one of the most widely cited questions around the globe, asked, 'Would you say that most people can be trusted, or that you need to be careful in dealing with people?' Too many of our pre-test respondents asserted, not unreasonably, that both propositions were true and they could not choose between them.

Where the survey questions are somewhat novel or phrased differently from other comparable surveys, this is clearly identified in relevant discussion of these results. The survey made specific effort at obtaining information on women's and children's work. Building on work done by the International Labour Organisation as well as time allocation studies done in India, special effort was made to determine women's and children's participation in caring for livestock, or in farm related activities. The resultant increase in netting women's work participation is discussed in detail in Chapter 4.

The questions finally fielded in IHDS were organized into two separate questionnaires, household and women. The household questionnaires were administered to the individual most knowledgeable about income and expenditure, frequently the male head of the household. The questionnaire for health and education was administered to a woman in the household, most often the spouse of the household head. Each interview required between 45 minutes and an hour- and-a-half to complete, a length that seemed the outer limits of what we could reasonably ask from our respondents. Questions on fertility, marriage, and gender relations in the households were addressed to an ever-married woman between 15–49 in the household. If no household member could fit the criteria, that portion of the questionnaire was skipped (about 19 per cent of all households). If the household had more than one ever-married woman between 15–49, one woman was selected randomly to answer those questions.

Because IHDS recognizes that all human development is nurtured within a local and institutional context, separate questionnaires were developed to measure village characteristics and to assess the functioning of up to two schools and two medical facilities located within the selected villages. In cases where there were no school and/or medical facilities within the selected village, the nearest school(s) and medical facility or facilities were surveyed. The data generated in the village, school, and medical facilities forms the basis of analysis carried out in Chapter 13.

FIELDWORK

The survey questions were originally drafted in English. However, given the multilingual diversity of India and large disparities in literacy levels, the questionnaires were then translated into Hindi for pre-testing, and then, after revisions, translated from the Hindi and English versions into 11 additional languages. The questionnaires translated in other languages were again pre-tested during training in the respective areas before these were used by the field teams to gather the information.

Fieldwork was performed by 25 agencies throughout the country, selected for their experience with administering large scale scientific surveys. A list of these collaborating organizations is included in Appendix II. The length and diversity of IHDS required more extensive training than is needed for single topic surveys. The NCAER staff, assisted by researchers from the University of Maryland, organized 11 two-week training sessions across the country, each for 15–50 interviewers. Classroom reviews of each questionnaire section alternated with supervised field experience. In addition to written interviewer manuals, training films were developed in which interviewers could see actual survey administration.

Once trained, interviewers went into the field typically in teams of five, two pairs of male and female interviewers and a team leader. The team leader was responsible for supervising and assisting with the household interviews and usually conducted the village, school, and medical facility interviews. After arriving at a PSU, the team would contact local leaders to describe the survey, secure permissions, and

develop a map of the area. Urban neighbourhoods and new villages selected in IHDS first required creation of a sampling frame. Large villages were divided into hamlets, or sections within the village, and two opposite sections were randomly selected for complete canvassing. Villages interviewed in the 1994 HDPI did not require canvassing and sampling, but the previous households had to be tracked, each member accounted for, and split households located.

Once the sample had been drawn or the 1993–4 HDPI households located, pairs of interviewers began arranging interviews. After obtaining consent, the household roster was filled out in duplicate. Separate households were defined as people living under one roof and sharing the same kitchen. Joint families often required specific probing since two married brothers might share the same dwelling but maintain separate kitchens and food budgets. Absent family members had to be identified as either temporarily absent household members (that is, living outside the household for less than six months), or residents of other households (for example, students living in nearby towns to pursue their education).

Once the household roster was completed, the two copies were divided between the two interviewers, and the female interviewer then completed the education and health questionnaire, usually with the help from a senior woman in the household. If the household included more than one eligible woman for the marriage and fertility sections, one was selected using a standard random number procedure. After completing the two main household interviews, the interviewers administered the learning tests to any child in the age group 8–11 years, and his/her height and weight measurements were taken. Often, more than one visit was needed to complete all sections of the household interview.

Completed interviews were checked by the team supervisor, rechecked by the agency coordinator and sent to NCAER headquarters in New Delhi, where editing staff again reviewed the skip patterns, looked for missing data, and checked coding. These multilevel reviews enabled prompt identification of problems and feedback to the interview teams. The NCAER also maintained its own field staff in each state for random re-interview checks for data quality and for troubleshooting of problems encountered by interview teams. Phone contact between agency field staff and NCAER headquarters also resolved many issues before they became major problems.

Data entry was centralized at NCAER's New Delhi offices and was undertaken as completed interviews arrived. The questionnaire form was mostly self-coded for ease of data entry. The 1,400 variables from the household interview were checked for consistency (for example, no five-year old mothers of three children) and problems resolved by consulting the originally filled questionnaire, or occasionally telephone calls back to the interview site. The main data files are publicly available for downloading and further analyses by all interested scholars. IHDS should become a premier resource for understanding the complexities of the human development process.

PUBLIC USE DATA

Data from IHDS 2005 are publicly available for free download from http://www.icpsr.umich.edu/cocoon/DSDR/STUDY/22626.xml. More information about the survey is available at www.ihds.umd.edu.

Appendix II—Chapter Organization and Definition of Variables

CHAPTER ORGANIZATION

Each of the chapters in this book reviews a major topic within the ambit of human development: income, education, health, social integration, and so on. Each topic has some issues specific to it (for example, marriage relations in Chapter 10). A few issues span more than one topic. For example, privatization is discussed in Chapters 4, 6, and 7. But as discussed in the introduction, a principal integrating theme is to review how these human development outcomes vary across a common set of social and economic determinants. Individual outcomes (for example, wages, employment, and morbidity) are compared along three characteristics of those individuals:

1. Gender
2. Age
3. Own education (adults only)

All individual and household outcomes are compared across five regional and household characteristics:

1. States
2. Rural–urban residence
3. Household income
4. Household educational level[1]
5. Religious and caste social group

The following section describes how each of these eight indicators is constructed, their distribution across India, and relationship with the other indicators. The sample distributions and interrelationships are presented in Table AII.1.

BACKGROUND CHARACTERISTICS

Gender

Each of the individual characteristics (that is, gender, age, and education) was reported by the main household respondent. This results in some imprecision pertaining to age and education, including the usual age heaping at round figure ages (20, 30, and so on). Some corrections have been made based on other information in the survey (for example, birth histories) but for comparisons of most human development outcomes, even imprecise measures are sufficient to reveal the strong patterns.

Measurement problems are not an issue for gender, although difficulties in locating transient and homeless populations may result in an undercount of men. India is well known for its imbalanced sex ratios and missing women. The IHDS also recorded fewer females than males, especially among the younger age groups, for whom the effects of sex selective abortions have become more apparent. The dynamics of gender inequality underlying these imbalanced

[1] Household educational level is used only for household level outcomes since individual outcomes are compared against the individual's own education.

Table AII.1 Sample Distribution Along Individual and Household Background Characteristics

	Rural	*Urban*	*Total*
Individual Characteristics			
Age			
0–4	10	8	9
5–9	12	10	11
10–14	12	11	12
15–19	10	11	10
20–9	16	19	17
30–9	13	15	14
40–59	18	20	18
60+	9	7	8
Sex			
Male	51	51	51
Female	49	49	49
Education			
Illiterate	44	26	39
1–4 Std	17	14	16
5–9 Std	27	30	28
10–11 Std	6	12	8
12 Some college	3	8	5
College graduate	2	10	4
Household Characteristics			
Social Group			
Forward Caste Hindu	16	31	21
OBC	38	31	36
Dalit	24	17	22
Adivasi	10	3	8
Muslim	10	14	11
Christian, Sikh, Jain	2	4	3
Place of Residence			
Metro		26	8
Other urban		74	21
More developed village	48		34
Less developed village	52		37
Maximum Adult Education in Household			
Illiterate	29	10	24
1–4 Std	10	5	8
5–9 Std	33	28	32

(*contd*)

(Table AII.1 *contd*)

	Rural	*Urban*	*Total*
10–11 Std	12	17	14
12 Some college	8	13	10
College graduate	8	27	13
Household Income			
Negative Rs 999	3	1	2
1st Quintile (Rs 1,000–14,000)	25	6	20
2nd Quintile (Rs 14,001–22,950)	23	10	19
3rd Quintile (Rs 22,951–36,097)	20	19	20
4th Quintile (Rs 36,098–69,000)	17	26	20
5th Quintile (Rs 69,001+)	12	38	19

Source: IHDS 2004–5 data.

sex ratios are examined in Chapters 8 and 10. Because women and men live in the same households, they don't differ greatly on household characteristics (although, somewhat more women live in low income households). Individual differences are substantial, however, as will be seen throughout the remaining chapters. Men average 5.1 years of education, for instance, compared to women's 3.6 years.

Age

India's fertility decline is fairly recent, so India is still a young country. Forty five per cent of IHDS household members are under 21. The young are somewhat more concentrated in poorer states, where the fertility decline has been the weakest, and in poorer households. There are more elderly (age 60 or more) in states with an early fertility decline (for example, Kerala), or where out-migration of the working age population leaves a higher concentration of the elderly (for example, Himachal Pradesh). Their well-being receives attention in Chapter 9. Age is inversely correlated with years of schooling since education has expanded manifold since independence. This correlation needs to be kept in mind in evaluating some tables since several human development outcomes tend to increase with both more education, and older ages.

Education

Education is one of the most consistent predictors of favourable human development outcomes. Everything from incomes to health to social connections is higher among the better educated. Because of educational expansion, India has many highly qualified graduates whose future is promising. The country also still has many illiterates whose struggles are often poorly rewarded. The tables that follow divide years of education into groups, according to the school system's natural break points. More than two in five adults have had no schooling. A small group, 9 per cent, started primary school without finishing. Over a quarter of adults finished primary school without completing secondary school. Almost a quarter of adults, however, have completed their 10th Standard. Ten per cent finished at that level, 6 per cent finished higher secondary school, and 7 per cent are college graduates.

Higher levels of education are more common among every advantaged group. Urban residents are more educated than rural residents. High income households have more educated members than poor households. Forward castes and non-Muslim minority religions have considerably more education, on an average, than other groups while Dalits and Adivasis have the least. Some of the many advantages of urban, affluent, forward castes result from their higher education, but some part of their higher education results from their many other advantages.

States

Regional inequalities have provoked a growing debate as parts of India have grown especially rapidly in recent years. Differences across states are a recurring theme in IHDS results, often overwhelming differences by class and social group. But there are limitations to the extent of state differences that can be reliably reported. The survey was fielded in thirty three states and union territories.

Sample sizes vary substantially across these states and territories (see Table AI.1). Care must always be taken not to rely too heavily on the position of any one state in the distribution of state outcomes. Sampling errors almost always overlap between states with similar positions on any

human development measure. Rather, much of the usefulness of state differences is to observe the pattern of state differences, rich versus poor, north versus south, and high versus low education.

While we report statewise results even for some samples that are quite small (for example, Uttarakhand), some of the union territories and states have samples too small to reliably report separate results. Therefore, these smaller samples had to be combined with neighbouring areas for reporting purposes (for example, Goa with Maharashtra). All the smaller north-eastern states (Arunachal Pradesh, Manipur, Meghalaya, Tripura, and so on) are reported as a single entity. These states share some common features, but are quite heterogeneous on many other dimensions. The other smaller states and territories were combined with larger neighbours, Chandigarh with Punjab, Daman and Diu, and Dadra and Nagar Haveli with Gujarat, Goa with Maharashtra, and, Pondicherry with Tamil Nadu. The Delhi sample is large enough to report separately for most purposes, but the rural sample in Delhi is based on only seven semi-urban villages, so Delhi is not reported separately for agricultural and other rural totals.

This organization leaves 22 'states' that are compared in each of the main chapters. For consistency, they are always reported in the same order rather than, for instance, from high to low on any outcome. Development is one common, but far from universal pattern distinguishing the 22 states. Urbanization, income, and education, is a coherent package that distinguishes states like Delhi and Kerala from states like Orissa and Jharkhand. There are exceptions, even within this development cluster, but it will be useful to think of this as one (among several) organizing principle for regional inequalities. However, other outcomes demonstrate quite a different pattern. Some social groups have strong state associations (for example, Muslims in Jammu and Kashmir; Christians in the North-East) but these are not usually development related (although tribal population more often reside in rural, less developed states and Sikhs in the wealthy Punjab). Some dimensions of gender inequality also cross-cut development levels. For example, unbalanced sex ratios are found in wealthy Punjab and poor Uttar Pradesh while more balanced sex ratios are found in affluent Kerala and poorer Orissa. The lesson here is that development levels are an important, but not the only dimension along which states in India differ.

Rural–Urban Residence

Village to city differences are a second type of regional inequality generally thought to be growing in recent years. Urban residents have higher incomes, their children stay in school longer, and when sick they have better access to medical care. While India has been slowly urbanizing throughout the last century, the pace of urbanization is only modest by world standards. In 2005, India had forty one urban areas with over a million population, while China had ninety five. Villages still hold much staying power, and even urban migrants maintain ties with their native villages. The perception of growing rural–urban disparities could threaten this stability.

The IHDS uses the Census 2001 definitions which classify as urban, places with a population of 5,000 or more and where most male employment is outside agriculture.[2] According to the 2001 Census, 28 per cent of India was urban. The IHDS slightly over sampled (34 per cent) urban areas but all analyses have been weighted back to the Census proportions.

Both urban and rural areas encompass great diversity. India's major metropolitan areas are the global cities. Mumbai's Bollywood is familiar to most of the world, Bangalore's IT industry, and Chennai's call centres daily influence the lives of millions of people outside India. At the other end of the spectrum, thousands of small towns are barely distinguishable from large villages. To capture these differences, IHDS reports urban results in two categories. The six largest metropolitan areas[3] (Mumbai, Kolkata, Delhi, Chennai, Hyderabad, and Bangalore) account for 7 per cent and all the other urban areas combined account for 21 per cent. Similarly, some villages have substantial infrastructure, paved roads with easy access to urban centres, postal and telephone connections, electricity to power lights, and televisions. Others lack most of the conveniences of modern life and can be reached only by narrow footpaths. In some cases one even has to use unconventional means like camel or boat. The IHDS divides villages into two approximately equal groups according to an index of infrastructural development described in the Chapter 12. The more developed villages generally appear closer to urban areas on most human development outcomes.

As discussed in Chapters 2 and 6, town, and especially metropolitan households, have higher incomes and education than rural households. This conflation of causal influences

[2] The official Census definition of an urban area is (a) All statutory places with a municipality, corporation, cantonment board, or notified town area committee, etc., or (b) A place satisfying the following three criteria simultaneously: i) a minimum population of 5,000, ii) at least 75 per cent of male working population engaged in non-agricultural pursuits, and iii) a density of population of at least 400 per sq. km (1,000 per sq. mile).

[3] The IHDS loosely follows the Census definitions of Urban Agglomeration which include areas outside the official municipal boundaries, but which are integrated into the urban core. All urban residents in districts identified as part of the urban agglomeration are counted as living in the metropolitan area. Census rules do not allow urban agglomerations to cross state boundaries, but we have included Gurgaon (Haryana), Ghaziabad, and Gautam Buddha Nagar (Uttar Pradesh) districts with the Delhi metropolitan area.

will often require that we look jointly at residence and socio-economic position in the chapters that follow to sort out which aspects of human development are specifically related to urbanization and which are a result of greater affluence and education. Urban areas also differ on their caste and religious composition. Forward castes and minority religions are especially concentrated in urban areas. Dalits and, especially, Adivasis are more rural.

Income

The IHDS is one of the first major Indian survey to measure detailed income. The NSS measures consumption expenditures and the *NFHS* measures the ownership of consumer goods. The IHDS measured these too. Each provides a somewhat different aspect of economic position, but is closely related. The details of their measurement and their inter-relationships are described in Chapters 2 and 5. The IHDS measure of income is summed across over 50 separate components including wages and salaries, net farm income, family business net income, property, and pension incomes.

The average Indian household had an annual income of Rs 27,857 in 2004. But because some households earned much more than this median, the mean was Rs 47,804. For all tables, households are divided into five quintiles with cutting points at 14,000, 22,950, 36,098, and 69,000. A small number of households (2 per cent) reported negative or very low incomes because of agricultural or business losses. Although these households are undergoing current economic distress, in many other ways (for example, consumer goods owned, educational levels, and so on) they appear more like moderate income households rather than poor households in the bottom quintile. They have, therefore, been excluded from the income tables, but are included in other analyses.

The income quintiles used throughout these reports do not vary across urban and rural areas, or across states, and, consequently, they do not adjust for price differences. Urban–rural price differences can be as large as 15 per cent.

Household Education

Many of the human development outcomes described in the previous chapters benefit the entire household. An indoor water tap, access to nearby medical clinics, and connections to government officials are resources the entire household can take advantage of. To see how these advantages are related to educational levels, the tables use a measure of the highest adult (that is, age 21 or older) education in the household, when appropriate.[4] The same schooling categories are used as for individual education, but the distribution is higher. Only a quarter of Indian households have no adult without any formal education, but 37 per cent have an adult who has matriculated, 10th Standard, or gone further. At the top, 13 per cent of households have an adult with a college degree. This measure of household education is associated with the same advantages as individual education. Urban residence, higher incomes, and forward castes are more common in well educated households. Note that the household educational attainment is greater than the individual one since household level education is based on highest education for any household member.

Social Groups

Perhaps no other country in the world offers such a rich diversity of religions, castes, ethnic, and linguistic identities, as it is found in India. Any useable grouping for a review of human development is bound to ignore important distinctions that the people themselves would never overlook. The tables here follow a six-fold classification:

1. Forward Castes
2. Other Backward Castes (OBC)
3. Dalits (Scheduled Castes)
4. Adivasis (Scheduled Tribes)
5. Muslims
6. Other Minority Religions (Christians, Sikhs, Buddhists, Jains)

The obvious question for such a scheme is where one classifies Muslim OBCs, Christian Adivasis, Sikh Dalits, and other groups, that easily fit more than one category. Muslim OBCs differ from Hindu OBCs and from other Muslims on most human development outcomes, and, likewise, for Christian Adivasis, Sikh Dalits, and other groups. Independent religion and caste classifications would avoid these ambiguities, but would create too many categories for the compact presentation needed here. The compromise result is this six category scheme described in Figure AII.1. More detailed classifications are available from the public data for analysts requiring more precision. Our construction of socio-religious categories has two major implications that must be kept in mind. First, 2,014 Muslim families, who classify themselves as OBCs form about 4.6 per cent of the total population, are included with Muslims rather than OBCs. Second, the inclusion of Christian, Sikh, and Buddhist Scheduled Caste families with Dalits and Adivasis, according to their self-classification, reduces the group classified as other minority religions from 6.29 per cent of the total population to 2.70 per cent (Figure AII.1).

[4] In households without any adult 21 years or older, the highest education is substituted.

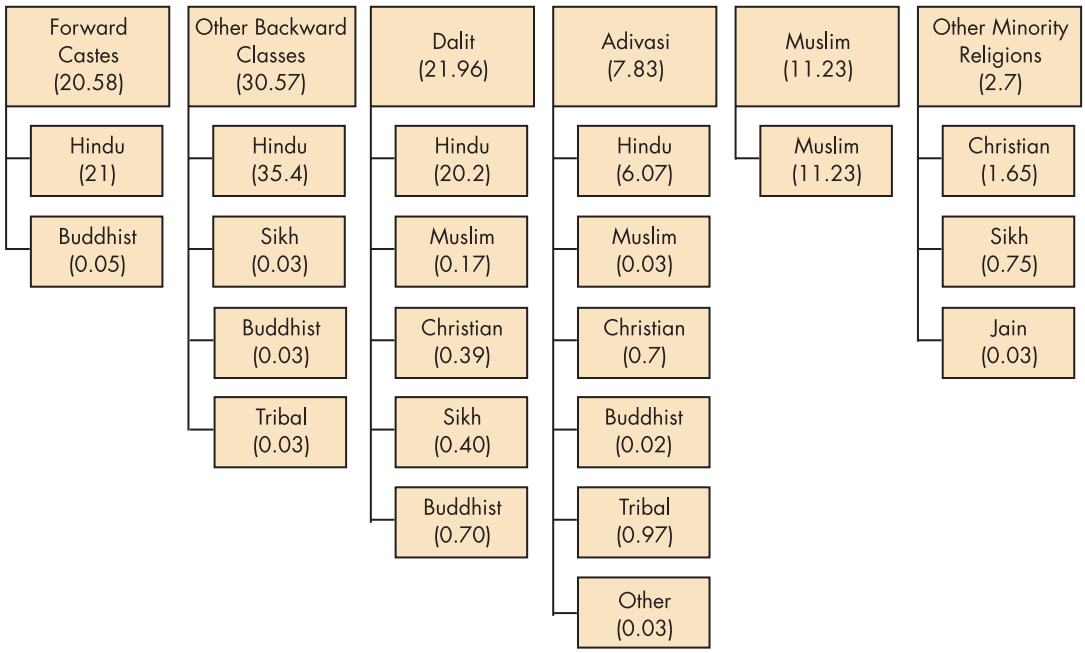

Figure AII.1 Socio-religious Group Categorization (in percentage)

Note: 276 households were selected as rural but became urban by 2001, bringing the total of urban households to 14,820.

Source: IHDS 2004–5 data.

Religion and caste classifications are based on the main respondent's self-identification. Self-identification yields somewhat different information from official data which use detailed but statewise government schedules. The official schedules often miss migrants from other states. Self-identification also encourages marginal groups to claim scheduled caste or tribe membership in order to qualify for government reservations. As a result the IHDS ends up with somewhat higher proportions of the population as Dalits and Adivasis than the Census figures, and slightly higher than the NSS.

The groups differ greatly on almost every measure of economic and social standing. Forward castes and non-Muslim minority religions are more urban, educated, and wealthy. Dalits and Adivasis are more often rural, illiterate, and poor. The OBCs are somewhere in between, but usually closer to Dalits than to forward castes. Muslims are also somewhere in between, but much closer to Dalits in education, closer to forward castes in urbanization, and in between on incomes, but slightly better off than the OBCs. These groups differ also on most of the human development outcomes we review in the previous chapters. Sometimes these differences are a result of the economic, educational, and regional differences, but sometimes some group differences remain even when comparing otherwise equivalent households.