

R

User's Guide for the Indonesia Family Life Survey, Wave 2

*E. Frankenberg, P. Hamilton, S. Polich,
W. Suriastini and D. Thomas*

DRU-2238/2-NIA/NICHD

March 2000

*Prepared for the National Institute on Aging/National Institute on
Child Health and Human Development*

Labor and Population Program

The RAND unrestricted draft series is intended to transmit preliminary results of RAND research. Unrestricted drafts have not been formally reviewed or edited. The views and conclusions expressed are tentative. A draft should not be cited or quoted without permission of the author, unless the preface grants such permission.

*RAND is a nonprofit institution that helps improve policy and decisionmaking through research and analysis.
RAND's publications and drafts do not necessarily reflect the opinions or policies of its research sponsors.*

We recommend the following citations for the IFLS data:

For papers using IFLS1 (1993):

Frankenberg, E. and L. Karoly. "The 1993 Indonesian Family Life Survey: Overview and Field Report." November, 1995. RAND. DRU-1195/1-NICHD/AID

For papers using IFLS2 (1997):

Frankenberg, E. and D. Thomas. "The Indonesia Family Life Survey (IFLS): Study Design and Results from Waves 1 and 2. DRU-2238/1-NIA/NICHD.

Preface

This document describes aspects of the public-use data from the Indonesia Family Life Survey, Wave 2 (IFLS2) to assist analysts in manipulating the data and constructing analytic files. It is the second of seven volumes documenting the IFLS2.

The Indonesia Family Life Survey is a continuing longitudinal socioeconomic and health survey. It is addressed to a sample representing about 83% of the Indonesian population living in 13 of the nation's 26 provinces. The survey collects data on individual respondents, their families, their households, the communities in which they live, and the health and education facilities they use. The first wave (IFLS1) was administered in 1993 to individuals living in 7,224 households. IFLS2 sought to reinterview the same respondents four years later. A follow-up survey (IFLS2+) was conducted in 1998 with 25% of the sample to measure the immediate impact of the economic and political crisis in Indonesia. The next wave, IFLS3, is scheduled to be fielded in 2000.

IFLS2 was a collaborative effort of RAND, UCLA, and the Demographic Institute of the University of Indonesia (LDUI). Funding for IFLS2 was provided by the National Institute on Aging (NIA), the National Institute for Child Health and Human Development (NICHD), U. S. Agency for International Development (USAID), The Futures Group (POLICY Project), the Hewlett Foundation, the International Food Policy Research Institute (IFPRI), John Snow International (the OMNI project), and the World Health Organization. MACRO International developed the data-entry software and had responsibility for some of the data processing.

The IFLS2 public-use file documentation, whose seven volumes are listed below, will be of interest to policymakers concerned about socioeconomic and health trends in nations like Indonesia, to researchers who are considering using or are already using the IFLS data, and to those studying the design and conduct of large-scale panel household and community surveys. Updates regarding the IFLS database subsequent to publication of these volumes will appear at the IFLS Web site, <http://www.rand.org/FLS/IFLS>.

Documentation for IFLS, Wave 2

DRU-2238/1-NIA/NICHD: *The Indonesia Family Life Survey (IFLS): Study Design and Results from Waves 1 and 2*. Purpose, design, fieldwork, and response rates for the survey, with an emphasis on wave 2; main results from both waves 1 and 2.

DRU-2238/2-NIA/NICHD: *Users Guide for the Indonesia Family Life Survey, Wave 2*. Descriptions of the IFLS file structure and data formats; guidelines for data use, with emphasis on using the wave 2 and wave 1 data together.

DRU-2238/3-NIA/NICHD: *Household Survey Questionnaire for the Indonesia Family Life Survey, Wave 2*. English translation of the questionnaires used for the household and individual interviews. Includes interviewer's instructions.

DRU-2238/4-NIA/NICHD: *Community-Facility Survey Questionnaire for the Indonesia Family Life Survey, Wave 2*. English translation of the questionnaires used for interviews with community leaders and facility representatives. Includes interviewer's instructions.

DRU-2238/5-NIA/NICHD: *Household Survey Codebook for the Indonesia Family Life Survey, Wave 2*. Descriptions of all variables from the IFLS2 Household Survey and their locations in the data files.

DRU-2238/6-NIA/NICHD: *Community-Facility Survey Codebook for the Indonesia Family Life Survey, Wave 2*. Descriptions of all variables from the IFLS2 Community-Facility Survey and their locations in the data files.

DRU-2238/7-NIA/NICHD: *Crosswalk between the Survey Instruments for the Indonesia Family Life Survey, Waves 1 and 2*.

Re-Release of IFLS1 Data

To facilitate using the IFLS1 and IFLS2 data together, a revised version of IFLS1 data has been released in 1999. Abbreviated IFLS1-RR (1999), the re-release incorporates adjustments outlined in the “fixes” files, joins subfiles having the same unit of observation, and adds identifiers that make it easier to link IFLS1 and IFLS2 data. The IFLS-RR data are available at <http://www.rand.org/FLS/IFLS> and are documented in

DRU-1195/7-NIA/NICHD: *Documentation for IFLS1-RR: Revised and Restructured Indonesia Family Life Survey Data, Wave 1*.

Previous Documentation for IFLS, Wave 1

DRU-1195/1-NIA/NICHD: *The 1993 Indonesian Family Life Survey: Overview and Field Report*. Purpose, design, fieldwork, and response rates.

DRU-1195/2-NIA/NICHD: *The 1993 Indonesian Family Life Survey: Appendix A, Household Questionnaires and Interviewer Manual*. English translation of the questionnaires used for the household and individual interviews. Includes interviewer’s instructions.

DRU-1195/3-NIA/NICHD: *The 1993 Indonesian Family Life Survey: Appendix B, Community-Facility Questionnaires and Interviewer Manual*. English translation of the questionnaires used for interviews with community leaders and facility representatives. Includes interviewer’s instructions.

DRU-1195/4-NIA/NICHD: *The 1993 Indonesian Family Life Survey: Appendix C, Household Codebook*. Descriptions of all variables from the Household Survey and their locations in the data files. Includes notes about cases that are known anomalies.

DRU-1195/5-NIA/NICHD: *The 1993 Indonesian Family Life Survey: Appendix D, Community-Facility Codebook*. Descriptions of all variables from the Community-Facility Survey and their locations in the data files. Includes notes about cases that are known anomalies.

DRU-1195/6-NIA/NICHD: *The 1993 Indonesian Family Life Survey: Appendix D, Users’ Guide*. Descriptions of the IFLS file structure and data formats; guidelines for data use, with emphasis on working with the household, individual, and facility IDs and making links across different parts of the survey.

Contents

Preface iii

Acknowledgments vii

1. Introduction 1

2. IFLS2 Data Elements Deriving from IFLS1 2

HHS: Reinterviewing IFLS1 Households and Individuals 2

HHS: Preprinted Household Roster 3

HHS: “Intended” Respondents and Households 4

HHS: Obtaining Retrospective Information 5

HHS: Updating Kinship Information 6

Siblings 6

Children 6

CFS: Reinterviewing IFLS1 Communities and Facilities 7

3. IFLS2 File Structure and Naming Conventions 11

Basic File Organization 11

Household Survey 11

Community-Facility Survey 11

Identifiers and Level of Observation 12

Household Survey 12

Community-Facility Survey 13

Combining Data across Files 14

Concatenating Data 14

One-to-one Merges at the Individual, Household, Community, or Facility Level 15

One-to-Many Merges 15

Merging HHS Data with CFS Data 16

Question Numbers and Variable Names 16

Response Types 17

Missing Values 18

Special Codes and X Variables 18

TYPE Variables 19

Privacy Protected Information 19

Weights 19

IFLS1 Household Weight 20

IFLS1 Person Weights 20

IFLS2 Weights 21

4. Special Features of the IFLS2 Data 30

Symmetric Information 30

Duplicate Information 30

Family Relationships 31

Parents, Children, and Spouses Identified in the AR Roster 31

Parents, Children, and Spouses Identified in Other Modules 32

Classifying Relatives 34

Identifying All of a Person's Closest Relatives 34

CFS: Using Information from Multiple Respondents 34

5. Cleaning the IFLS Data 37

In the Field: CAFÉ Editing, Interviewer Rechecks 37

In Jakarta 37

Double Data Entry and Verification 37

"Look Ups" 38

Special Cleaning for Open-ended, "Other," and Numeric Variables 39

In Santa Monica 40

Module Checks 40

Checks on IDs across Books and Survey Waves 41

Checks on Book Covers 41

Checks on Preprinted Child and Sibling Rosters 41

Checks on Units of Measure 41

Created Variables and Files 42

6. Using IFLS2 Data with IFLS1 Data 43

IFLS1 Re-Release 43

Differing IFLS1 and IFLS2 Household IDs 43

Merging IFLS1 and IFLS2 Data for Households and Individuals 44

Data Availability for Households and Individuals (HTRACK and PTRACK) 45

HTRACK 45

PTRACK 45

Tracking Changes in Characteristics across Survey Waves 47

Data Availability for Communities and Facilities: CTRACK and FTRACK 47

Merging IFLS1 and IFLS2 Data for Communities and Facilities 48

Appendix

A: Names of Data Files for the Household Survey 50

B: Names of Data Files for the Community-Facility Survey 54

C: Module-Specific Analytic Notes 57

D: Special Cases 68

Glossary [70]

Acknowledgments

A survey of the magnitude of IFLS2 is a huge undertaking. It involved a large team of people from both the United States and Indonesia. We are indebted to every member of the team. We are grateful to each of our respondents, who gave up many hours of their time.

The project was directed by Elizabeth Frankenberg (RAND) and Duncan Thomas (RAND and UCLA), who were the Principal Investigators. Lynn Karoly and Paul Gertler were Principal Investigators in the early stages of the project.

Bondan Sikoki was the Project Director appointed by the Demographic Institute of the University of Indonesia (LDUI). She served as the Survey Director during the design and implementation of fieldwork. Her unswerving commitment to maintaining the integrity and quality of IFLS2, in even the most difficult circumstances, was an inspiration to us all. Prior to her appointment, the LDUI Project Director was Dr. IGN Agung.

Three LDUI staff members served as Associate Project Directors. Wayan Suriastini directed the tracking phase of the study and played a central role in the design of the Household Survey Questionnaire. Muda Saputra coordinated much of the Community-Facility Survey fieldwork and data entry. Sutji Rochani Siregar oversaw the administration of the latter phases of fieldwork and data entry.

Data-entry software and field procedures for the Computer-Assisted Field Editing (CAFE) were developed by Trevor Croft, of MACRO International, with the assistance of Hendratno of LDUI. Croft also developed the software used for the final phase of data entry/data quality checks (Look Ups). Iip Umar Ri'fai, Martin Wolfe, and Linda Fitrawati assisted with these tasks.

Eko Ganiarto coordinated the first and second pretests. Victoria Beard worked extensively on the Community-Facility Survey. Endjang Pudjani and Sheila Evans were responsible for the technical production of the Indonesian and English questionnaires. Akhir Matua Harahap coordinated the writing and production of the survey manuals. Mary Linehan managed operations in Jakarta prior to fieldwork; she developed the assessments of physical health, along with Cecep Sukria Sumantri and Merry Widayanti. Nargis, Djainal, and M. Yusuf assisted with the development of the Community-Facility Survey and the training of its staff. Donavan Bustami coordinated printing and shipping for the questionnaires.

John Adams provided critical input for the design of the follow-up protocols and guided the development of sampling weights. Christine Peterson designed the preprinted rosters, assisted with questionnaire design and processing of the pretest data, and helped calculate the sampling weights.

The IFLS2 public-use data files were produced by a team based at RAND. The efforts of Paula Hamilton, Nancy Campbell, Melissa Chiu, Sue Polich, Patty St. Clair, Wayan Suriastini, and Peter Yau went well beyond the call of duty.

Many of our colleagues at RAND have contributed substantially to the survey. We are especially grateful to James P. Smith and John Strauss. We are also grateful to Kathleen Beegle, Julie DaVanzo, William Dow, Micki Fujisaki, Doug Gilbertson, Paul Gertler, Daryl Hill, Michael Hurd, Lynn Karoly, Jacob Klerman, Nancy Krantz, Donna Lee, Lee Lillard, Maria Menchaca, Eileen Miech, Jack Molyneaux, Mathew Sanders, Christine d'Arc Taylor, Jim Tebow, and Beverly Weidmer.

Much effort was put into designing IFLS2 so that it would yield information on topics of special concern in Indonesia and reflect the nation's distinctive social, economic, and policy environment. The input of a large number of scholars and policy-makers in Indonesia was key in this regard. Paramita Sudharto gave us considerable guidance on the overall survey and on its health components. Important contributions were made by Boediono, Mark Brook, Fasli Djalal, Herwindo Haribowo, Bachrul Hayat, Heryudarini, Yayah Husaini, Bambang Indrianto, Stephanus Indradjaya, Jiono, Robert Kim-Farley, Vanda Moriaga, Dr. Mujilah, Muljani Nurhadi, Ratna, Kusnadi Setjawinata, Soeharsono Soemantri, James Stein, Ace Suryadi, and Anton Wijaya.

The survey could not have taken place without the support of the LDUI directors and administrative staff, including N. Haidy Pasay, Sri Moertiningsih Adioetomo, Sri Hariati Hatmadji, Badrun, and Teguh. We are indebted to the Population Study Centers in each of the thirteen IFLS provinces, which helped us recruit the 400 field staff.

Finally, the success of the survey is largely a reflection of the diligence, persistence and commitment to quality of the interviewers, supervisors, and field coordinators. Their names are listed in the *Study Design* (DRU-2238/1-NIA/NICHD), Appendix A.

1. Introduction

The Indonesia Family Life Survey is rich but complex. This guide discusses aspects of the IFLS data to assist analysts in manipulating the data and constructing analytic files. Information on sample design, recontact rates, sample sizes, and questionnaire content is provided in the *Study Design* volume (DRU-2238/1-NIA/NICHD), which also presents analytic results on selected topics.

The second wave of the IFLS (IFLS2) was fielded in 1997, four years after the first wave. Because the IFLS is a panel survey, many elements of IFLS2 are based on IFLS1. Section 2 of this guide describes how the IFLS2 built on IFLS1 with respect to sample composition and the types of data collected. Section 3 describes the file structures and conventions used in the data, including how files and variables were named, identifiers, types of variables, and codes used to indicate missing data. This section also explains the weights that are available for use with the data.

Section 4 explains some special features of the IFLS, with emphasis on ways the data can be used to identify family relationships. Multiple data modules contain information on relationships among parents, children, siblings, and spouses. The various information sources are described, with suggestions on how to combine data to yield the most complete picture of family ties.

Throughout the process of collecting the data and preparing the public-use files, we implemented a variety of procedures to maintain a high level of data quality. They are described in Section 5.

Finally, Sec. 6 describes how to use the IFLS2 data in combination with IFLS1. To simplify their joint use, we have issued a revised version of the IFLS1 data called the IFLS1 Re-Release, or IFLS1-RR (1999). Section 6 provides guidelines for using the IFLS1-RR as well as files we have constructed to provide summary information for all individuals (PTRACK), households (HTRACK), communities (CTRACK), and facilities (FTRACK) that were interviewed in either IFLS1 or IFLS2. We also describe how to merge IFLS1 and IFLS2 data for individuals, households, communities, and facilities.

Appendixes A and B list the names of electronic data files provide for the Household Survey and Community-Facility Survey, respectively. Appendix C provides detailed notes of analytic interest about particular data modules. They include comments on data collection strategy or question content that affect the comparability of IFLS2 and IFLS1 data, problems observed in the field or during data cleaning, and warnings about mistakes to avoid in using the data. Appendix D provides a list of “special cases,” variables or records with unique characteristics that could not be reflected in the electronic data. Analysts may want to handle these variables and records differently from others of their type.

2. IFLS2 Data Elements Deriving from IFLS1

This section discusses elements of the IFLS2 data that derive from IFLS1. The bulk of the discussion applies to the Household Survey (HHS), with the Community Facility-Survey (CFS) covered at the end of the section.

HHS: Reinterviewing IFLS1 Households and Individuals

As explained in Sec. 2 of the *Study Design* (DRU-2238/1-NIA/NICHD), IFLS2 attempted to reinterview all 7,224 households interviewed in IFLS1. For each of those *panel households*,¹ a preprinted roster was generated. It listed the household's IFLS1 ID and the name, age, sex, birthdate, and relationship to the household head of all members of the household in 1993.

Interviewers were instructed to return to the household's 1993 address. If none of the 1993 members was still in residence, the interviewers were instructed to look for them. To assist field staff in finding panel households, a relocation sheet was preprinted for each household with detailed information from IFLS1: the household's address and the name, age, and gender of every household member. For target respondents additional detail included places of employment and schools; place of birth; all places of residence; and names of non-coresident family members, including parents, siblings and children. Finally, the sheet listed information we had the foresight to ask in IFLS1: where respondents might go if they were to leave, and the name of a person in the current area who might know their whereabouts in a few years.

At the point of first contact with any 1993 household member, the original household was said to have been found. An interview was conducted under the same household ID, with current information collected for everyone listed in the preprinted roster. As a result, in the vast majority of cases an *origin household* resided

At the household's 1993 location and included most of the 1993 members,

but other scenarios also occurred, where the origin household resided

At a distant location from the 1993 residence but with the household intact

At a different location with a few 1993 household members

At the same location but with very few of the 1993 household members.

We also sought interviews with households that had "split-off" from panel households. They were defined as households containing a *target respondent*—either an IFLS1 household member who had provided detailed individual-level information in 1993 or who had been 26 or older in 1993.

¹ Italicized terms and acronyms are defined in the Glossary.

Application of the “first contact” rule for an origin household² sometimes yielded odd results. Hypothetical examples:

In a 1993 household of 5 people, all had moved from the 1993 location by 1997. The 17-year-old son was living next door with his aunt so that he could finish his schooling. The others had moved far away. Since the son was the first to be contacted, his was designated the origin household. When traced to their new location, the four other original members were designated a new split-off household. It might seem more intuitive to call the four members who remained together the origin household and the son with his aunt’s family the split-off household, but the rule dictated otherwise.

Only a servant was found remaining in the 1993 location. In that origin household, everyone else was recorded as having left the household, the servant’s new employer was designated the household head, and the relationship of all the former members to the current household head was designated “non-relative.”

One way of spotting such anomalies in origin households is to look for households that have a large number of people listed in the roster, with high proportions of 1993 members who have left (AR01a = 3), a high proportion of new members (AR01a = 5), and a small number of remaining members (AR01a = 1). In using IFLS2 data generally, remember that not all individuals listed in the household roster for origin households were current members of the household in 1997.

Another apparent anomaly is that for a small number of households (around 80), a household roster exists but includes no current members (AR01a never = 1, 4, or 5). This occurred either because all the 1993 household members had died by the time the interview team arrived in the EA, or because the only 1993 household members still alive in 1997 had joined another IFLS household by the time of the 1997 interview.

HHS: Preprinted Household Roster

In certain modules, information collected in IFLS1 was preprinted on survey forms and used in IFLS2 interviews. The purpose was twofold: to ensure that information on particular households and individuals was updated and to save time during the interview. To avoid the associated disadvantages,³ we limited the use of preprinted material to modules that required lists of names, where updating was essential and the potential for saving time the greatest.

The most important example of preprinted information (others are discussed later in this section) was the preprinted household roster. For every panel household, a roster was generated that contained the following information for each IFLS1 household member:

² We established the first-contact rule because it was the best way of ensuring that at least some information was gathered for all IFLS1 household members. Postponing use of the preprinted household roster until the “most logical” origin household was found would have risked losing altogether the opportunity for a comprehensive accounting by a 1993 household member of the whereabouts of the other 1993 members.

³ Using preprinted material requires that the field team be well organized and pay attention to detail to get the correct preprinted forms to the correct households. Also, errors in the preprinted information can confuse interviewers and respondents.

Person Identifier in 1993

Name

Sex

Age

Birthdate

Relation to the household head in 1993

Tracking status (whether the person was a target respondent)

Panel status for books 3 and 4 (whether the person gave detailed information for IFLS1 book 3 or 4)

When an origin household was found, the interviewer inserted the household's preprinted roster as the base page in book K, and the interviewer asked for updated information about each member on the list. Occasionally the preprinted roster contained the name of someone listed as a household member in 1993 whom the 1997 respondent had never heard of. Occasionally the preprinted roster did not list someone who the 1997 respondent said had been living in the household in 1993. Special response categories for AR18f (reason for entry into/exit from household) were created to identify these cases.

The preprinted roster was invaluable in making sure that IFLS2 collected at least some information about every 1993 household member. When a target respondent had moved out of the household, his or her preprinted information was transferred onto a tracking form that was used to collect information about where the person had gone.

For split-off households we used a blank sheet rather than preprinted roster as the base page in book K. All members of the new household were manually listed on the page. *PIDLINKs* (defined in Sec. 6) and panel status information were transferred from the tracking forms onto the base page for individuals who had been tracked from the origin household to the new household.⁴

HHS: "Intended" Respondents and Households

In IFLS2 we sought to reinterview all IFLS1 households and split-off households that contained a target respondent. For obtaining household-level information, interviewers were asked to administer books K, 1, and 2 to a household member 18 or older who was knowledgeable about household affairs. Generally book 1 was answered by a female (usually the female household head) and book 2 was answered by a male (usually the male household head). However, these were guidelines, not strict rules. A household book was sometimes answered by someone outside the household, usually when the household members were too sick or disabled (for example, hard of hearing) to give the information. In that case, the respondent was often a relative or caregiver. Occasionally a household book was answered by someone younger than 18 because he or she was the most knowledgeable person available. The covers of books K, 1, and 2 provided space to record the identifier of the person answering the book and that person's relationship to the household head.

With respect to individuals, in IFLS2 we sought to interview all current members of an origin household. In split-off households, we sought to interview the target respondent, his or her spouse, and all biological

⁴ All split-off households contained at least one person who was a member of an IFLS1 household. In split-off households, AR01a = 4 for individuals who were tracked to that household from an IFLS1 household. For all other members of the split-off household, AR01a = 5, indicating that they were new to IFLS.

children. For obtaining individual-level information, the books administered depended on whether the person was a panel respondent and on his or her age, sex, and marital status.

Respondents age 15 and older were supposed to answer books 3A and 3B, and respondents under age 15 were supposed to answer book 5. For IFLS1 household members, preprinted information indicated whether the person should answer books 3A and 3B or book 5. If a respondent was expected to be 15 or older by 1997, he or she was supposed to be administered books 3A and 3B. In the field, interviewers sometimes encountered respondents who said they were younger than 15 but the preprinted information indicated that they were 15 or older. Rather than override the preprinted instructions, interviewers generally administered both books 3A and 3B and book 5.

Information about children and pregnancies was collected in both books 3B and 4. For IFLS1 women respondents, preprinted information indicated which of those books the woman should answer. If she had answered book 4 in 1993, she was asked to answer it in 1997. This protocol meant that some women who answered book 4 in 1997 were in their early 50s (whereas book 4 was technically limited to women 15–49). If a woman had not answered book 4 in 1993, she was asked to answer it in 1997 if she was between the age 15 and 49 and was currently married or had previously been married.

Book 5 was administered to all household members younger than age 15. Children 11–14 were allowed to answer for themselves; an adult (usually the mother) answered for children younger than age 11.

Inevitably we were not successful at administering all indicated books to all intended households and individuals. Sometimes we could not find a household or respondent. In other cases households or individuals were found but respondents refused to be interviewed.

Anticipating the impossibility of interviewing all the respondents from whom we wanted information, we designed a proxy book to obtain a subset of information from someone who could answer for a respondent. The proxy book contained many of the modules from books 3A, 3B, and 4, but most modules asked for less information than the “main” books. For example, we collected data about only two of a woman’s pregnancies. The proxy book also provided a “Don’t Know” option more frequently than the main books. The person who completed the proxy book was usually someone who knew the respondent well, such as the respondent’s spouse or parent.

Table 2.1 indicates the differences in information obtained from the proxy book and corresponding main books of the survey.⁵ To make full use of the available individual-level information, the analyst should append data from the proxy book to the related data from books 3A, 3B, and 4.

To help analysts identify which respondents provided data for which books, we created files named PTRACK and HTRACK. They indicate who answered what and provide codes regarding nonresponse for individuals and households, respectively.⁶

HHS: Obtaining Retrospective Information

A number of modules in books 3A, 3B, and 4 were designed to collect retrospective information from respondents. Examples are modules on education, marriage, migration, labor force participation, pregnancies, and contraceptive use.

⁵ In this document, numbered tables appear at the end of the section where first cited.

⁶ These files are described in more detail in Sec 6.

Respondents who had provided detailed information in IFLS1 (i.e., panel respondents) were not asked to provide full histories again in IFLS2. For respondents who had not answered Books 3A, 3B, or 4 in IFLS1, it was necessary to request the “full” history.

The covers of books 3A, 3B, and 4 provided a place to record each respondent’s panel status for that book, as indicated on the preprinted household roster. In addition, modules that collected retrospective information usually contained a “panel check” whereby the interviewer ascertained whether the respondent was panel or new and followed a different skip pattern depending on the answer.

IFLS2 generally collected less information about panel respondents than about new respondents. The questionnaires were structured (1) to collect the same retrospective information for new respondents as had been collected in IFLS1 and (2) for panel respondents, only to update the information collected in IFLS1 with information about what had happened since. Therefore, to provide full retrospective information for IFLS2 panel respondents, the analyst must link data from both waves. To facilitate linking, the IFLS2 collected some information about events or behavior in 1992 or 1993, providing an overlap between what was reported in both waves. For certain modules, additional data were collected from panel respondents to permit assessments of the quality of the retrospective reporting.

Table 2.2 summarizes the differences in information collected from new and panel respondents in the retrospective modules and their implications for creating a full history for panel respondents.

HHS: Updating Kinship Information

In IFLS1 certain respondents were asked very detailed information about their siblings and children. Rather than burdening respondents with the time-consuming task of relisting those relatives in IFLS2, we preprinted rosters of siblings and children for interviewers to use.

Siblings

In IFLS1, book 3 respondents were asked about all non-coresident siblings age 15 or older who were alive or who had died within the previous 12 months. In IFLS2, to save time for respondents who had reported such siblings in 1993, the names of all *living* non-coresident siblings from IFLS1 were listed in a preprinted roster. IFLS2 respondents to book 3B who did not have a preprinted roster (e.g., a new respondent or panel respondent who had reported no qualifying siblings in 1993) filled out a sibling roster from scratch.

Children

In IFLS1, women respondents to book 3, module BA, were asked to list all non-coresident children, including any who had died within the previous 12 months. Men respondents were asked to complete module BA and list those children only if their wife was not a household member or if they had had children with women other than the wife currently in the household. Other IFLS1 modules collected information on children, e.g., the household roster (module AR) and the pregnancy history (module CH).

To reduce the burden for IFLS2 respondents, we created preprinted child rosters for respondents who had provided information on their children in IFLS1 and thus were expected to be eligible for the BA module in IFLS2. Rather than limiting the rosters to children not residing in the household in 1993, we listed all living children identified by the 1993 respondent. In addition to the children’s names, we listed their line numbers from any IFLS1 module in which they were listed (AR, BA, or CH). Because of the selection rules for providing child information in 1993, a woman was much more likely than a man to have a preprinted child roster in 1997.

IFLS2 respondents who did not provide child information in 1993 (so did not have a preprinted child roster) but were eligible to do so in 1997 completed a BA child roster from scratch. That group included men whose wife was no longer a household member, women who had answered book 3 or book 4 in 1993 but who had no children at that time, and new respondents.

The administration and associated data processing of the preprinted sibling and child rosters were among the most complicated elements of IFLS2. Analysts are urged to read the comments about module BA in Appendix C.

CFS: Reinterviewing IFLS1 Facilities and Communities

Whereas a primary goal of the HHS was to reinterview households and individuals interviewed in IFLS1, the CFS aimed at describing the communities and available facilities for households and individuals interviewed in IFLS2. We sought to maintain comparability with the IFLS1 CFS instruments, but we were not explicitly trying to obtain high recontact rates for facilities or respondents interviewed in 1993.

At the community level the CFS in both IFLS1 and IFLS2 sought interviews with two officers of the community: the head of the community and the head of the women's group. To the extent that there was continuity in the holders of those positions, the same individuals were interviewed in both waves. For community-level information, we have not attempted to determine whether particular respondents in 1997 were also respondents in 1993.

With respect to facilities, the same sample selection procedure was used in IFLS2 as in IFLS1. To the extent that there was little turnover in the facilities available to respondents, and IFLS1 interviewed a high fraction of the available facilities, many of the facilities interviewed in 1993 were interviewed again in 1997.

To assist in matching facilities across waves, we had panel facilities assigned the same ID in both years.⁷ In the field, reassignment of the 1993 ID to a facility was accomplished with the Service Availability Roster (SAR). The roster included a preprinted list of the names, addresses, and IDs of facilities mentioned in IFLS1 as being available within the EA. Completing the SAR required (1) noting whether each facility on the preprinted list was still available in 1997 and (2) listing any facility newly available to community members since IFLS1 that was identified by either an HHS respondent or community informant. In using the SAR to finalize the facility sampling list, the field supervisor assigned the 1993 ID to any facility noted as still being available in 1997.

Unlike the HHS, which collected much retrospective information from respondents, the CFS collected relatively little retrospective information. In book 1 for community leaders, only one module asked about community history. In IFLS1 community leaders were asked about major community-level events going back to 1980. In IFLS2, the leaders were asked only about events going back to 1992.

⁷ The exception is community health posts. No community health post interviewed in IFLS2 has the same ID as its IFLS1 counterpart. That is because both the locations and volunteer staff changed over time, so determining whether an IFLS2 post was the same as an IFLS1 post was effectively impossible. It is perhaps more appropriate to regard a community health post as an activity rather than a facility.

Table 2.1
Differences in Information Collected from Proxy Book vs.
Corresponding Main-Book Module

Module	Information in Proxy Book	Additional Information in Main Book
KW	Current marital status Dowry, residence decisions associated with current or most recent marriage	Date started co-residing and information on who else was in the household History of marriages Fertility preferences
MG	Birthplace, residence at age 12, date of move to current residence and place from which respondent moved	History of migrations
DL	Literacy, educational level, date of school completion (or departure), grade repetition, EBTANAS scores, expenditures on schooling in previous year	Characteristics of schooling at each level attended (elementary, junior high school, senior high school, post-secondary)
TK	Current work status, date and earnings from last job if not currently working, hours and wages of current primary and secondary jobs, date of first job	History of jobs over the last five to nine years
PM	Participation in an <i>arisan</i> , whether borrowed money, participation in community development activities	Detail on arisan participation, knowledge and use of credit institutions, levels and forms of participation in community development activities
KM	Whether ever smoked, what was smoked, and length of time since quitting (if not a current smoker)	Detail on quantity smoked
KK	Health conditions – no difference in proxy and main-book information	
MA	Experience of morbidity in past month	Chest pain, injuries that were slow to heal
RJ	Incidence and reasons for visits to health care providers in the past 4 weeks	Detail on services received and expenditures on care
RN	Inpatient visits – no difference in proxy and main-book information	
BR	Children living outside the household, children that died, stillbirths, and miscarriages	Number of children in the household
CH	Pregnancy outcome, use of prenatal care, delivery site, survival status for up to two pregnancies	Detail on prenatal services received, length of labor, birthweight, breastfeeding
BA	Non-coresident family and transfers – no difference in proxy and main-book information	

Table 2.2
Differences in Information Collected from New vs. Panel Respondents in IFLS2

Module	New Respondents	Panel Respondents	Creating a Full History for Panel Respondents
DL (education) Panel check: DL07x	Highest level of education attained and on each level of schooling attended.	<p>Highest level of schooling attended since 1991 for</p> <ul style="list-style-type: none"> • panel respondents still attending school at IFLS2 • panel respondents younger than age 25 at IFLS2 who had attended school since 1991 <p>Note: panel respondents younger than 50 who had not attended school since 1991 were not asked their highest level of educational attainment (this information is available in IFLS1 and in module AR)</p>	<p>Use data from IFLS1 module DL for schooling before 1993.</p> <p>Schooling between 1992 and 1993 is reported in both IFLS1 and IFLS2</p>
DLR (schooling disruptions)	All disruptions of schooling in the past 5 years. This module was new in IFLS2, so the same information was collected from panel and new respondents.		
KW (marriage) Panel check: KW22x	All previous and current marriages	Current or most recent marriage and any other marriage that began after 1991	For respondents who have had no marriages that ended before 1993, IFLS2 provides a complete marriage history. Data on marriages that ended before 1993 are in IFLS1.
MG (migration) Panel check: MG00x	Residence at birth, age 12, and all moves after age 12	All moves after age 12	Use IFLS1 for birthplace and residence at age 12.
TK (employment) Panel check: TK47x	Primary and secondary jobs for a 5-year period (back to 1992)	Primary and secondary jobs for a 10-year period (back to 1987)	IFLS2 contains more information on panel respondents than on new respondents. For panel respondents, additional information is available in IFLS1 on employment in 1983 and in 1973.

Table 2.2 (cont.)

3A, BR (pregnancy summary) Panel check: BR00x	All live births, still births, and miscarriages (for new respondents at least age 50)	None	Use IFLS1 for pregnancy summaries for panel respondents who were 50 or older at IFLS1.
4, BR (pregnancy summary) Panel check: BR00x	All live births, still births, and miscarriages (new respondents and panel respondents without a preprinted child roster)	None if panel respondent had preprinted child roster	Use IFLS1 for births up to 1993. Use IFLS2 data in the CH module to compute the number of additional births since 1993.
BF (breastfeeding) Panel check: BF00	Asked in module CH (new respondents and panel respondents without a preprinted child roster)	Update on breastfeeding for the youngest child at the 1993 interview if that child was 8 or younger in 1997 (therefore might still have been breastfeeding in 1993)	If the youngest child was still breastfeeding in 1993, use IFLS2 data in BF00 to determine the total duration of breastfeeding. For children born since 1993, breastfeeding data are in IFLS2.
CH (pregnancies) Panel check: CH00	All pregnancies (new respondents and panel respondents without a preprinted child roster)	Pregnancies occurring after the birth of the child who was the youngest child in 1993 (panel respondents with a preprinted child roster) Note: for panel respondents to book 4 who had a preprinted roster, information on the total number of pregnancies or children ever born cannot be calculated without using IFLS1	Use the IFLS1 data in the CH module for pregnancies that began before 1993. Some pregnancies that occurred in 1992 and 1993 may be in both IFLS1 and IFLS2.
KL (contraceptive use) Panel check: KL00x	Contraceptive use patterns beginning in January 1992 or date of first marriage, whichever was later	Contraceptive use beginning in <ul style="list-style-type: none"> January 1987 or date of first marriage (whichever was later) for a random subset of 25% of panel respondents January 1992 for the other 75% of panel respondents 	The IFLS2 calendar contains at least one year of overlap with the IFLS1 calendar for panel respondents. For the 25% subsample there is a longer period of overlap.

3. IFLS2 File Structure and Naming Conventions

This section describes the organization, naming conventions, and other distinctive features of the IFLS2 data files to facilitate their use in analysis. Additional information about the data files is provided in the survey questionnaires and codebooks. For analysts' convenience, each page of the HHS and CFS questionnaires includes the names of files that contain information from that page. The codebook for each questionnaire book describes the files containing the data for that book and the levels of observation represented.

Basic File Organization

Files containing HHS and CFS data are available in ASCII, SAS, and Stata formats.

Household Survey

HHS data files correspond to questionnaire books and modules. There are multiple data files for a single questionnaire module if the module collected data at multiple levels of observation. For example, module DL (education history) collected information at the individual level (on educational attainment) and at the school level (on characteristics of schools the respondent attended at each level), so two data files are associated with that module.

File naming conventions are straightforward. The first two or three characters identify the associated questionnaire book, followed by characters identifying the specific module and a number denoting sequence if data from the module are spread across multiple data files:

Bxx_xxx	x
Book_Module	File sequence

Continuing the above example, the name B3A_DL1 signifies that the data file contains information from book 3A, module DL, and is the first of multiple files. The name B3A_DL2 denotes the second file of information from book 3A, module DL. Appendix A lists the name of each data file from the HHS, along with the associated level of observation and number of records.

Community-Facility Survey

CFS data typically have one file at the community or the facility level that contains basic characteristics and spans multiple questionnaire modules within a book. Additional files at other levels of observation are included when appropriate, as explained below.

Files are named by the questionnaire book. Data files that contain information at the level of the community or the facility are named as follows:

CFS Book	Corresponding File Name	CFS Book	Corresponding File Name
Book 1	BK1	Book PUSK	PUSK
Book 2	BK2	Book PP	PRA
Book PKK	PKK	Book Posyandu	POS
Book SAR	SAR	Book SD	SD
Book Adat	ADAT	Book SMP	SMP
Book PM	PM	Book SMU	SMU

Names of additional files containing information at another level of observation also identify the associated module, e.g., BK1_A. For example, consider book 1, module A. The first page has a grid that repeats several questions (e.g., travel time) for various institutions or destinations. This information is included in file BK1_A, in which each observation is an institution or destination. Module A also contained questions such as whether the community offers a public transportation system and the prevailing price of gasoline. For these questions, there is one answer for each community, so the answers are in file BK1. File BK1 also contains community-level data from other modules such as whether the community has piped water or a sewage system. Appendix B lists the name of each data file from the CFS, along with the associated level of observation and number of records

Identifiers and Level of Observation

Household Survey

Wherever possible the data have been organized so that the level of observation within a file is either the household or the individual. If the level of observation is the household, variable HHID97 uniquely identifies an observation. If the level of observation is the individual, both HHID97 and PID97 are required to uniquely identify a person.⁸

In IFLS2, HHID97 is a seven digit character variable whose digits carry the following meanings:

x	x	x	x	x	x	x
EA			specific household		origin/split-off	

In the last two digits, 00 designates an origin household. For a split-off household, the 6th digit is always 1, signifying a split in IFLS2, and the 7th digit indicates whether it is the first, second, or other split-off (some multiple split-offs occurred).

In IFLS2, the person identifier PID97 is simply the line number of the person in the AR roster.

⁸ Within IFLS2 files, use HHID97 and PID97 to identify individuals. In the IFLS2 AR roster, variable PIDLINK does not uniquely identify individuals because individuals can be listed in more than one household roster (but they are a current member of only one household – see Sec. 6).

When the level of observation is something other than the household or individual, it is usually because¹³ the data were collected as part of a grid, in which a set of questions was repeated for a series of items or events. For example, in the health care provider data from module PP, each observation corresponds to a particular type of provider, and there are multiple observations per household. In this data file, the combination of HHID97 and PPTYPE uniquely identifies an observation. The variable that defines the items or events is usually named XXXTYPE, where XXX identifies the associated module (more is said about TYPE variables below).

In some cases, data collected as part of a grid are organized rectangularly. For example, file B1_PP1 contains data about 11 provider types for each of 7,566 households. Thus, there are $11 \times 7566 = 83,226$ observations in the data file. In other cases, the number of records per household or individual varies. For example, the level of observation in file B3B_RJ is a visit by an individual to an outpatient provider. Not all individuals made the same number of visits, so some individuals appear only once, others appear twice, and some appear more than twice. Those who made no visits do not appear at all. This file is not rectangular because the number of observations per person is not constant. To uniquely identify an observation in this file, the analyst should use HHID97, PID97, and RJTYPE.

Community-Facility Survey

Wherever possible, CFS data are organized so that the level of observation within a file is either the community or the facility. In a community-level file, an observation can be uniquely identified with COMMID97. In a facility-level file, an observation can be uniquely identified with variable FCODE.

The first two digits of variable COMMID97 identify the province, and the remaining two digits indicate a sequence number within the province:

XX	XX
Province	Sequence

The following codes identify the 13 IFLS provinces:

12 = North Sumatra	34 = Yogyakarta
13 = West Sumatra	35 = East Java
16 = South Sumatra	51 = Bali
18 = Lampung	52 = West Nusa Tenggara
31 = Jakarta	63 = South Kalimantan
32 = West Java	73 = South Sulawesi
33 = Central Java	

The first four digits of variable FCODE are COMMID97, the fifth digit indicates the facility type, and the last two digits indicate the facility type's sequence number within the community.⁹

⁹ FCODEs did not change between 1993 and 1997, and some facilities were used by members of more than one IFLS community. Note that the community ID embedded in FCODE is not necessarily the community in which the facility is located, or the community for which the facility was interviewed, or the only IFLS community to which the

<u>x x x x</u>	<u>x</u>	<u>x x</u>
COMMID97	Facility type	Sequence

The codes for facility type are the following:

- 1 = health center or subcenter (*puskesmas* or *puskesmas pembantu*)
- 2 = private practitioner (*praktek* or *klinik swasta*, *praktek* or *klinik umum*)
- 3 = private practitioner (*praktek* or *klinik swasta*, *praktek* or *klinik umum*)
- 4 = community health post (*posyandu*)
- 5 = traditional practitioner (e.g., *dukun*, *sinse*, *tabib*, *tukang pijat*)
- 6 = elementary school (*sekolah dasar* or SD)
- 7 = junior high school (SMP, SLTP)
- 8 = senior high school (SMA, SMU)

Data were sometimes collected as part of a grid (defined above), such as types of equipment in health facilities or types of credit institutions in a village. The items or events are usually defined by a variable named XXXTYPE, where XXX identifies the associated module. The data in grids are rectangular where the number of observations per community or facility is fixed and are not rectangular where the number of observations varies. To uniquely identify an observation within a grid, use either COMMID97 or FCODE (if the data are from a facility questionnaire) and XXXTYPE for that data file. For the SAR, it is necessary to use both COMMID97 and FCODE to uniquely identify an observation because some facilities were shared by multiple communities, so an FCODE may appear more than once in the SAR.

Combining Data Across Files

As explained above, IFLS data are stored in many different data files. To create analytic files, the analyst usually needs to combine the data from different files. How the data should be combined depends on the nature of the desired analytic file. Below we briefly describe ways to link data across files.

Concatenating Data

The analyst may wish to pool observations by concatenating two data files. For example, B3B_RJ2 and B5_RJA2 both contain data on visits to outpatient providers. The data in B3B_RJ2 pertain to adults, and the data in B5_RJA2 pertain to children. The variables for adults begin with RJ, while the variables for children begin with RJA, but otherwise the information is the same. In some contexts it may be useful to combine the data for the two age groups, rather than keeping it in two separate files. The data can be combined into one file using the APPEND statement in STATA or the SET statement in SAS. The resulting file will contain both the observations for children and the observations for adults. Because the variable names are different, the variables in one file should be renamed so that they match the names in the other file.

facility provides services. To identify which facilities provide services to an IFLS community, analysts should use the Service Availability Roster (SAR). See Sec. 4, “CFS: Using Information from Multiple Respondents.”

Many files could conceivably be concatenated. Here we address one combination that is particularly important. As a general rule, when using data from books 3A, 3B, and 4, check whether a corresponding module was included in the proxy book, so that data from respondents who answered for themselves can be combined with data collected by proxy for other individuals (see Sec. 2 for more about proxy information). 15

Table 3.1 lists additional combinations. Some files will need to be restructured before they are concatenated to account for differing levels of observation. Some files will need to have variables renamed.

One-to-One Merges at the Individual, Household, Community, or Facility Level

In many cases the analyst will want to link data from one file with data about the same respondents from another file. If both files contain data at the same level of observation, the linkage will be a “one-to-one” merge.

Merging Two Files at the Individual Level of Observation. Suppose the goal is to create a file that contains information on an individual’s literacy and his or her primary activity in the past week. The file B3A_DL1 contains information on whether respondents can read or write. The file B3A_TK1 contains information on the respondents’ primary activity in the past week. Both files contain one observation per individual. To link the desired information, sort each of the two files by HHID97 and PID97 and then merge on HHID97 and PID97.

Merging Two Files at the Household Level of Observation. For two household-level files, such as B2_KR and BK_KR, which contain data on housing characteristics, sort each file by HHID97 and merge on HHID97.

Merging Two Files at the Community Level of Observation. Generally it is not necessary to merge two files at the community level, because for each type of community respondent we have pooled all the community-level information into one file (see the preceding section). The analyst may want to combine community-level information collected from different respondents, since although the respondents are different, they are referring to the same community. An example is BK1 and PKK. In this case, sort each file by COMMID97 and merge on COMMID97. In this case it will be necessary to rename variables in either BK1 or PKK, because some variables have the same name in each file. If variables were not renamed so that each name appears in only one file, merging the files would overwrite data from one of the files.

Merging Two Files at the Facility Level of Observation. The variable FCODE uniquely identifies a facility. However, it should not be necessary to merge two files at the facility level, because for each type of facility we have pooled all the facility-level information into one file (see the preceding section).

One-to-Many Merges

Often the analyst will want to merge files that are not organized at the same level of observation. Sometimes such a merge is straightforward. Other times it will require restructuring at least one of the data sets. When thinking about how to merge IFLS data files, it is helpful to determine whether the identifying variables in one of the files are a subset of the identifying variables in the other file.

This is easiest to explain using an example. Suppose the analyst wishes to merge information on literacy with information on asset ownership. The identifying variables in B3A_DL1 are HHID97 and PID97. The

identifying variables in B3A_HR1 are HHID97, PID97, and HRTYPE. In B3A_HR1, an individual has 16 11 records, one for each asset type about which we inquire. The data can be merged in two ways.

First, because the identifying variables in B3A_DL1 are a subset of the identifying variables in B3A_HR1, you could simply merge on HHID97 and PID97. This yields 11 records for each individual. Each record contains information about the individual's literacy and information about a particular asset type.

The other option is to restructure B3A_HR1 so that it is organized at the level of the individual rather than at the level of the asset, and the identifying variables are HHID97 and PID97. This would involve creating a file that contained variables HR01–HR12 for asset type A (i.e., HR01A–HR12A), as well as variables HR01–HR12 for the other asset types (HR01B–HR12B, HR01C–HR12C, etc.). This file would have many more variables than B3A_HR1 but many fewer observations. If the data from the B3A_HR1 file are restructured to be at the level of the individual, merging the restructured file by HHID97 and PID97 with the B3A_DL1 data yields one record per person that contains literacy information and all information on the different types of assets.

Restructuring data files so that they are organized at a different level of observation can be done relatively easily in STATA with the reshape commands, or in SAS with PROC TRANSPOSE.

Some data files cannot be merged without restructuring one of the data files. For example, the identifying variables in B2_UT2 are HHID97 and UTTYPE. The identifying variables in B2_NT2 are HHID97 and NTTYPE. Neither file's identifying variables are a subset of the other's identifying variables. To merge data from these two files, you must first restructure one or both of them so that HHID97 is the identifying variable. Generally, it is not wise to merge two files that both contain data from grids (and have a TYPE identifying variable) without restructuring the data.

Merging HHS Data with CFS Data

HHS data files are never organized at the same level of observation as data files in the CFS, but generally it is not necessary to restructure the data unless the community or facility data are from a questionnaire module that contains a grid.

To merge HHS data with CFS data from community-level respondents, use variable COMMID97, which can be found in HTRACK. COMMID97 must be merged with the household- or individual-level file before that file can be merged with the community-level data. An individual or household matches to data from no more than one community – that of his or her current residence. Some individuals (those who no longer live in IFLS communities) will not match to any community data.

To merge HHS data with CFS data from facilities, use variable FCODE, which is found in both CFS facility data and in HHS data from questionnaire modules asking respondents to identify specific facilities that they knew about or used, i.e., PP, RJ and RJA, RN and RNA, DL and DLA, CH, and KL. A particular individual may be associated with (and thus match to) multiple facilities. For example, a woman may have used one facility for an outpatient visit and another facility for contraceptive supplies. In that case, she will have one FCODE in her RJ record and another FCODE in her KL record.

Note that individuals or households that had moved out of IFLS enumeration areas by 1997 and were interviewed elsewhere will not merge to any community or facility data because community and facility data were not collected in the communities to which people moved. Also, among individuals who did

not move, some will not merge to any facility data because they did not use facilities that were interviewed as part of the facility sample.

17

Questions Numbers and Variable Names

Most IFLS2 variable names closely correspond to survey question numbers. For example, the names of variables from the DL module (education history) begin with DL and end with the specific question number.

In the IFLS2 questionnaire we tried to number the questions so as to preserve the correspondence with IFLS1 question numbers. If a question was added or changed in IFLS2, we typically added “a” or “b” to the question number rather than renumbering questions and destroying the correspondence. For example, the first three questions of module DL in IFLS2 preserve the correspondence with related questions in IFLS1 while the addition of “a” for two numbers signifies subtle differences in content:

IFLS1		IFLS2	
DL01	Is Indonesian used at home?	DL01a	What languages are used at home?
DL02	Can the respondent read an Indonesian newspaper?	DL02	Can the respondent read an Indonesian newspaper?
DL03	Can the respondent write a letter in Indonesian?	DL02a	Can the respondent read a newspaper in another language?

For a module-by-module crosswalk between the questions in IFLS1 and the questions in IFLS2, see the *Crosswalk* (DRU-2238/7-NIA/NICHD).

A number of questions have two associated variables: an X variable indicating whether the respondent could answer the question and the “main” variable providing the respondent’s answer. X variables are named by adding “x” to the associated question number. For example, question DL07b asked when the respondent stopped attending school. Variable DL07bx indicates whether the respondent was able to answer the question. Variable DL07b provides the date school attendance stopped. In the questionnaire, the existence of an x variable is signaled when the interviewer is asked to circle a number indicating whether the respondent was able to answer the question (in the case of DL07bx, 1 if a valid date is provided, 8 if the respondent doesn’t know the date). In the codebooks, the name of the variable itself signals its X status. The label for an X variable includes an “able ans” at the end. X variables are further discussed below.

Response Types

The vast majority of IFLS questions required either a number or a closed-ended categorical response; a few questions allowed an open-ended response.

The numeric questions generally specified the maximum number of digits and decimal places allowed in an answer; any response not fitting the specification was assigned a *special code* by the interviewer, and the special codes were reviewed and recoded later (explained further below). Where it was necessary to add digits or decimal places as a result of that review, we may not have updated the questionnaire. The codebook provides information on the length of each variable.

Questions requiring categorical responses usually allowed only one answer (for example, Was the school you attended public or private?). When only one answer was allowed, numeric response codes were specified. If more than four numeric response codes were possible, two digits were used so that 95–99 could serve as special codes. Some questions allowed multiple answers (for example, What languages do you speak at home?). In that case, alphabetic response codes were specified. When multiple responses were allowed, the number of possible responses set the maximum possible length for the variable. 18

For categorical variables, the questionnaire provides the full meanings for each response category. The codebook contains a short “format” that summarizes the response category, but analysts should check the questionnaire for the clearest explanation of response categories and not rely solely on the codebook format.

The codebook also provides information on the distribution of responses. For numeric variables, the mean, maximum, and minimum values are given. For categorical variables the frequency distribution is provided. For categorical variables where multiple responses were allowed, the codebook provides the number of respondents who gave each response. Since many combinations of responses were possible, the codebook does not provide the distribution of all responses. For example, question DL01a asked what languages the respondent used in daily life and allowed up to 22 languages in response. The codebook shows how many respondents cited Indonesian and how many respondents cited Javanese but not how many respondents cited both Indonesian and Javanese.

Additional response categories were sometimes added in the process of cleaning “other” variables (discussed in Sec. 5). Typically these categories were added below the existing “other” category. For example, question DL11 asked about the administration of the school. The questionnaire as fielded provided six substantive choices and a seventh, “other.” When the “other” responses were reviewed, an eighth category, “Private Buddhist,” was added.

Missing Values

Missing values are usually indicated by special codes. For numeric variables, a 9 or a period signifies missing data. For character variables, a “z” or a blank signifies missing data.

For many variables, we can distinguish between *system missing data* (data properly absent because of skip patterns in the questionnaire) and data missing because of interviewer error. The data entry software generated some missing values automatically as a result of skip patterns. For example, question HR00a in book 3A asked the interviewer to check whether the respondent already answered module HR in book 2, and if so, to skip to the next module. If the interviewer recorded 1 (Yes), during data entry the software automatically skipped to the next module and filled the book 3A HR variables with a period or blank. If data were missing because the interviewer neglected to ask the question or fill in the response, the data-entry editor was forced to enter 9 or z in the data fields in order to get to the questions that the interviewer did ask.

Sometimes valid answers are missing not because of skip patterns or interviewer error but because the answer did not fit in the space provided, the question was not applicable to the respondent, the respondent refused to answer the question, or the respondent did not know the answer. In these cases special codes ending in 5, 6, 7, or 8 were used rather than 9 or z (see below).

Special Codes and X Variables

19

Many IFLS2 questions called for numeric answers. Sometimes a respondent did not know the answer or refused to answer. Sometimes the respondent said that the question was not applicable. Sometimes the answer would not fit the space provided, either because there were too many digits or decimal places were needed. Sometimes the answer was missing for an unknown reason. In all of these cases, interviewers used special codes to indicate that the question had not been answered properly. The last digit of a special code was a number between 5 and 9, indicating the reason:

5 = out of range, answer does not fit available space

6 = question is not applicable

7 = respondent refused to answer

8 = respondent did not know the answer

9 = answer is missing

The other spaces for the answer were filled with 9's so that the special code occupied the maximum number of digits allowed.

Rather than leave special codes in the data, we created indicator (X) variables showing whether or not valid numeric data were provided. An indicator variable has the same name as the variable containing the numeric data except that it ends in X. For example, the indicator variable for PP7 (expected price of services at a certain facility) is PP7X. The value of PP7X is 1 if the respondent provided a valid numeric answer and 8 if the respondent did not know what to expect in terms of prices.

An indicator variable sometimes reveals more than whether special codes were used. For example, for PP5 (travel time to a certain facility), PP5X indicates both the units in which travel time was recorded (minutes, hours, or days) and the existence of valid numeric data. Similarly, for PP6 (cost of traveling to the facility), PP6X indicates whether the respondent gave a price (= 1), walked to the facility (= 3), used his or her own transportation (= 5), or didn't know the answer (= 8).

For questions asking respondents to identify a location, X variables are used to indicate whether the location was in the same administrative area as the respondent (= 3) or a different administrative area (= 1). These X variables are typically available at the level of the *desa*, *kecamatan*, *kabupaten*, and province. For example, PP4aX indicates whether the facility identified by the respondent is located in the respondent's village or a different village.

TYPE Variables

As noted above, in some modules the data are arranged in grids, and the level of observation is something other than the household or individual. Examples are KS (household expenditure) data on prices, where the level of observation is a food or non-food item; PP (outpatient care) data, where the level of observation is a type of facility; and TK (employment) data, where the level of observation is a year. The name of the variable that identifies the particular observation level typically contains the module plus "TYPE," e.g., PPTYPE. In modules with TYPE variables, there are multiple records per household or individual, but combining HHID or HHID and PID with the TYPE variables uniquely identifies an observation. TYPE data can be either numeric or character.

In compliance with regulations governing the appropriate treatment of human subjects, information that could be used to identify respondents in the IFLS survey has been suppressed. This includes respondents' names and residence locations and the names and physical locations of the facilities that respondents used. Translations of open-ended responses do not include information that might help identify respondents.

Weights

The IFLS sample, which covers 13 provinces, is intended to be representative of 83% of the Indonesian population in 1993. By design, the original survey over-sampled urban households and households in provinces other than Java. It is therefore necessary to weight the sample in order to obtain estimates that represent the underlying population. This section discusses the IFLS1 and IFLS2 sampling weights that have been constructed for use with the household data. An overview of the weights is provided in Table 3.2.

IFLS1 Household weights

When the household weights that were included with IFLS1 are applied to the sample, the resulting weighted distribution will reflect the 1993 distribution of households in rural and urban areas within each of the 13 provinces covered by the IFLS. Those weights are the inverse of the sample selection probabilities for each household interviewed in IFLS1 so, intuitively, a household with weight ω can be viewed as representing $\omega\pi$ households in the underlying population where π is the number of households in the population (181,548,000) divided by the number of households interviewed in IFLS1 (7,224). Sample design effects are summarized in Table 3.3¹⁰

IFLS1 Person weights

Person weights in IFLS1 are complicated by the within-household sampling scheme adopted in the first round of the survey. There are three individual weights.

¹⁰The household weights were constructed by comparing the distribution of IFLS households, stratified by province and urban-rural sector with an estimate of the distribution of all Indonesian households stratified in the same way. The population estimate of households in each cell was obtained by dividing BPS' projected population for 1993 by the average household size in the 1993 SUSENAS. Two weights were released with IFLS1. HHWT224 which is the weight that should be applied to all 7,224 households that were successfully interviewed in 1993. We will refer to those as the 1993 household weights (HWT93). They are included in the household tracking data file, HTRACK. (Weight variables have been renamed so that a common convention can be used through all waves of IFLS.) The IFLS1 household weight HHWT730 is calculated including all 7,730 households that were in the original IFLS1 sampling frame; that frame included more households than the target sample of 7,000 in anticipation of a 10% incompleteness rate.

Roster weights are assigned to every individual listed in the IFLS1 household roster. The weights are designed so that the weighted age and sex distribution of individuals in IFLS1 reflect the 1993 population age and sex distribution by urban and rural strata within the 13 provinces covered by the survey.¹¹

IFLS1 *respondent weights* take into account the within-household sampling scheme used to select respondents for individual interviews in IFLS1. IFLS1 conducted detailed interviews with the following household members:

- The household head and spouse;
- Two randomly selected children of the head and spouse aged 0 to 14 (interviewed by proxy);
- An individual age 50 and above and their spouse, randomly selected from the remaining members;
- For a randomly selected 25 percent of households, an individual age 15 to 49 and his or her spouse, also randomly selected from remaining members.

We refer to the selected respondents as the IFLS1 *Main Respondents*; if their responses are weighted by the respondent weight, they should be representative of the underlying population.¹²

IFLS1 *anthropometry weights* take into account the within-household sampling scheme used to select the respondents who were weighed and measured. All IFLS1 *Main Respondents* along with all children under age 6 living in the household were eligible for anthropometric measurement.¹³ Anthropometric indicators that are weighted using the *anthropometry weights* will be representative of all Indonesians in the 13 IFLS provinces.¹⁴

¹¹The roster weight is based on all household members listed in the roster (Book 1, Section AR). The data were stratified by province, urban-rural sector, sex and five-year age groups (except for individuals age 75 and older who were treated as one group). The proportions in each strata were matched to the population proportions estimated from the 1993 SUSENAS. The person-level weight for 1993, PWT93, is recorded in the person-level tracker file, PTRACK. It is the same as ROSTERWT in IFLS1.

¹²Using the selection rules, the probability a member of a household would be selected was computed given all household members listed in the roster. That probability was inverted, normed and capped (at 3 which is the 99 percentile of the weight distribution) to obtain the IFLS1 respondent weight. It is called PWT93IN and is in PTRACK; the variable was called RESPWT in IFLS1.

¹³IFLS1 Main Respondents who were measured were given an anthropometry weight equal to their respondent weight (unnormed and uncapped); other children under age 6 were given the household weight (based on the 7,224 household sample). Household members who were measured but not eligible (i.e., they did not fit the selection criteria) were given an anthropometry weight of zero. The initial anthropometry weight was then normalized to sum to the number of those across all households who were eligible to be measured, to account for the fact that not all household members eligible for anthropometric measurement were actually measured. Finally, as with the respondent weight, the anthropometry weight was capped at 3 to control for those with very small probabilities of selection.

¹⁴The anthropometry weights, PWT93US, are in PTRACK, they were called CA_WT in IFLS1.

There are two types of weights for IFLS2 respondents. *IFLS2 longitudinal analysis weights* are intended to update the IFLS1 weights because of attrition so that the IFLS2 panel sample is representative of the Indonesian population living in the 13 IFLS provinces in 1993. All respondents who were interviewed in 1997 but were not in an IFLS1 household roster are *new entrants* in IFLS2; they are assigned a longitudinal analysis weight of zero. It might be argued that the full sample of respondents interviewed in IFLS2 is sufficiently similar to the Indonesian population living in Indonesia in 1997 that one could use the sample to describe that population. Since the IFLS1 sample design included over-sampling in urban areas and off Java, users will need to re-weight the sample to take these design effects into account. The *IFLS2 cross-section analysis weights* are intended to do just that.

IFLS2 longitudinal analysis household weights

If all IFLS1 households were re-interviewed in IFLS2, the IFLS1 household weights and IFLS2 longitudinal analysis household weights would be identical. The IFLS2 longitudinal analysis household therefore comprise two conceptually distinct components:

- Sample design effects that are embodied in the IFLS1 household weight, HWT93 (called HHWT224 in IFLS1).
- An adjustment for household-level attrition between IFLS1 and IFLS2.

Fortunately, household-level recontact rates in IFLS2 are high, relative to other household surveys conducted in the United States and in developing countries. An interview was conducted with at least one member of an IFLS1 household in 93.5% of cases; in 0.9% of cases, all IFLS1 household members had died by the time of IFLS2 leaving 5.6% of households who could have been interviewed but were not. See the *Overview* for detail and Thomas, Frankenberg and Smith (1999) for a fuller discussion of attrition in IFLS.

Low attrition rates notwithstanding, adjusting for attrition is controversial. It involves model-building and necessarily incorporates judgments that may not be appropriate for some analyses. In those cases, users should rely on the 1993 weight, HWT93, or derive and apply their own attrition adjustments.

Attrition in a panel survey is the outcome of interactions among a complex set of factors including the characteristics of the underlying population, the sample respondents and the survey design and operation. (See, for example, Little and Rubin, 1987, and Groves and Couper, 1998, for discussions.) Recognizing this, our goal is to provide some general purpose weights for analysis of the IFLS data. We have therefore adopted a simple model of between-wave attrition that we think captures the key characteristics of those households that were not re-interviewed in IFLS2. Taking a propensity score approach to constructing the weights, we estimated a logit model of the probability a respondent was found in IFLS2,¹⁵ computed the predicted probability the household was found and inverted that

¹⁵To be precise, we estimated (1 - the probability a household that could be interviewed was not interviewed).

probability to obtain an implied attrition adjustment for each household.¹⁶ Estimates from the logit model are reported in Table 3.4. 23

The attrition adjustments were capped at the 99th percentile. The product of the capped attrition adjustments and the IFLS1 household weights which incorporate sample design effects yield a household weight for each IFLS1 household that was found in IFLS2. We refer to this weight as ω_{HH1} .

The design of IFLS2 called for following all *target respondents* -- all IFLS1 *Main Respondents* and all other IFLS1 household members who were born prior to 1968 -- if that person had moved out of the household by the time of the IFLS2 interview. Those target respondents who had moved generated *split-off* households and so a single IFLS1 household can spawn multiple IFLS2 households. The IFLS2 household weights take this into account by distributing the estimated weight for the original household, ω_{HH1} among the IFLS2 households. Specifically, assume κ IFLS1 household members were re-located in IFLS2; each of those IFLS2 respondents is assigned $(1/\kappa)$ of the weight ω_{HH1} associated with their origin household. Taking the sum of these individual-assigned weights yields the IFLS2 longitudinal analysis household weight, HWT97L.

As an example, say there were 3 people in the original IFLS1 household; 2 were found in the origin location and 1 had split off; that respondent was found in a new location in a household with 1 other person. The attrition adjusted household weight, ω_{HH1} , is split equally among the three original household members who were found and so the origin household is assigned a weight of $\frac{2}{3}\omega_{HH1}$ and the split-off household is assigned a weight of $\frac{1}{3}\omega_{HH1}$. The new entrant (to the survey) in the split-off household does not enter the calculation. There are a small number of cases in which members of two different IFLS1 households combined into a single IFLS2 household. In those instances, the calculation of the IFLS2 longitudinal analysis household weight follows the same principle and is the sum of individual-assigned weights based on the IFLS2 respondents origin household in IFLS1.

Analyses of IFLS2 household data should use HWT97L to obtain estimates that are weighted to reflect the Indonesian population in the 13 IFLS provinces in 1993. Analyses that recombine IFLS2 households so that they match one-to-one with IFLS1 households should add up the weights, HWT97L, associated with these households and use the sum of the weights in the estimation.

IFLS2 longitudinal analysis person weights

The IFLS2 longitudinal analysis person weights follow a similar approach. A logit model of attrition was estimated for all individuals in the IFLS1 household rosters; the model excludes all new entrants in IFLS2. The inverse of the predicted probability of being considered as completed in IFLS2,¹⁷ conditional on being in an IFLS1 roster, yields the attrition adjustments. Models of attrition were estimated separately for *target respondents* and all other respondents because only target respondents were followed if they had moved out of the household and so the probability of re-contacting them -- and the characteristics associated with that probability -- is different from the other respondents. Estimates from the logit models are reported in Table 3.5.

¹⁶Households in which all members of the IFLS1 households had died by 1997 are treated as found in these calculations.

¹⁷An individual is considered completed if the respondent was found in an IFLS2 household or is known to have died between the waves.

The individual-specific attrition adjustments were capped at the 99th percentile of the relevant distribution of weights for *target respondents* and other respondents and multiplied by the IFLS1 household weights to take into account sample design effects. The result is PWT97L, the IFLS2 longitudinal analysis person weight variable, which is recorded in PTRACK, the person level tracking file. PWT97L is set to 0 for all individuals in IFLS2 who were not listed in an IFLS1 household roster. Estimates that are weighted with this variable should correspond with the 1993 Indonesian population in the 13 IFLS provinces.

The same procedure was followed to construct longitudinal analysis person weights for use with the anthropometric measures. In IFLS1, a sub-sample of respondents were weighed and measured. In IFLS2, we sought to conduct physical health assessments on all respondents; the completion rate was around 85% of all IFLS2 respondents. Analyses that exploit the repeated measures of heights and weights in IFLS1 and IFLS2 will generate estimates that are representative of the 1993 population if weighted by the health assessment person level longitudinal analysis weights in 1997, PWT97USL.

IFLS2 cross-section analysis person weights

While IFLS is a longitudinal survey, there will be some analyses that only use information collected in IFLS2 because, for example, comparable data was not collected in IFLS1. Some analyses will, therefore, effectively treat IFLS2 as if it were a cross-section. We have attempted to construct weights so that estimates based on IFLS2 will be representative of the Indonesian population living in the 13 IFLS provinces at the time of IFLS2.

It is not obvious how to do this. After some experimentation, we have followed a procedure that parallels the approach taken to construct roster weights in IFLS1 and raked the IFLS2 sample (after adjusting for attrition) to an external sample, the 1997 wave of the SUSENAS which is thought to provide a good representation of the Indonesian population at that time.

All individuals listed as being present in the IFLS2 households have been stratified by province and urban-rural sector of residence, by sex and by age (into 5 year age groups with everyone 75 and above in a single group). These cell proportions have been reweighted using the attrition adjustments calculated from the individual-specific logistic regressions in Table 3.5 and then matched to the cell proportions in the 1997 SUSENAS. The *IFLS2 cross-section analysis person weights* are the ratio of the SUSENAS proportion to the IFLS2 proportion in each cell. The resulting weights are called PWT97X and are included in PTRACK.

Estimates that use these weights should be representative of the Indonesian population in 1997 in the IFLS provinces. In view of the timing of IFLS2 (August 1997-February 1998), one could argue that IFLS2 should be raked to the 1998 SUSENAS (conducted in February 1998). We have chosen to not do that because February 1998 was a time of tremendous turmoil in Indonesia and a time when it was thought that large numbers of people were re-locating (in part in response to incentives from the Indonesian government to "return to the *desa*".)

Similar weights have been constructed for use with the health assessments, PWT97USX, and the cognitive assessments, PWT97EKX. These weights were constructed by raking to the 1997 SUSENAS and take into account the fact that the assessments were not completed by all eligible respondents.

IFLS2 cross-section analysis household weights

An analogous strategy has been adopted to construct cross-section analysis weights at the household level. All households in the IFLS2 sample have been stratified by province and urban-rural sector; the cell proportions have been weighted by the attrition adjustments implied by the household-level logistic regression reported in Table 3.4. For each cell, the ratio of the proportion of households in the 1997 SUSENAS sample to the weighted proportion of IFLS2 households provides the *IFLS2 cross-section analysis household weights*, HWT97X, which are included in HTRACK. Estimates that are weighted with HWT97X should be representative of all households living in the IFLS provinces in Indonesia in 1997. 25

Table 3.1
HHS Files Suitable for Concatenation

Topic	Files	Respondent Types
Assets	B2_HR1 and B3A_HR1	Book 2 and book 3A respondents
EBTANAS scores	B3A_DL3 and B5_DLA1	Young adults and children
Schooling behavior	B3A_DL3 and B5_DLA3	Adults and children
Schooling disruptions	B3A_DLR2 and B5_DLA2	Young adults and children
Marriage	B3A_KW1 and B4_KW1 B3A_KW3 and B4_KW3	Ever-married women 14-49 and other adults
Pregnancy summary	B3A_BR & B4_BR	New female respondents age 50 and older and new female respondents younger than 50
Outpatient visit summary	B3B_RJ1 and B5_RJA1	Adults and children
Outpatient visit detail	B3B_RJ2 and B5_RJA2	Adults and children
Inpatient visit summary	B3B_RN1 and B5_RNA1	Adults and children
Inpatient visit detail	B3B_RN2 and B5_RNA2	Adults and children
Non-coresident children	B3B_BA6 and B4_BA6 and B4_CH1	Ever-married women 15-49 and other adults

Table 3.2: Sample design effects in IFLS1

Province	Census Code	Indonesian population			Susenas 1993		IFLS 1993		
		000s HHs	% HHs	% Urban	# EAs	Sampling rate	# EAs	# Urban	# Rural
North Sumatra	12	10,391	5.7	35	732	2:1	26	16	10
West Sumatra	13	4,041	2.2	20	502	3:1	14	6	8
South Sumatra	16	6,403	3.5	29	428	2:1	15	8	7
Lampung	18	6,108	3.4	12	244	2:1	11	3	8
DKI Jakarta	31	8,352	4.6	100	380	2:1	40	40	0
West Java	32	35,973	19.8	33	1282	1:1	52	31	21
Central Java	33	28,733	15.8	26	1578	1:1	37	19	18
DI Yogyakarta	34	2,923	1.6	48	216	4:1	22	16	6
East Java	35	32,713	18.0	26	1814	1:1	45	23	22
Bali	51	2,798	1.5	27	320	4:1	14	7	7
West Nusa Tenggara	52	3,416	1.9	17	244	4:1	16	6	10
South Kalimantan	63	2,636	1.5	23	380	4:1	13	6	7
South Sulawesi	73	7,045	3.9	24	912	2:1	16	8	8
TOTAL		181,548	100.0		9,032		321	189	132

Table 3.3: Summary of weights

Original Name	IFLS1 WEIGHTS	IFLS2 WEIGHTS		
	Re-release Name	Longitudinal Analysis	Cross-Section Analysis	
HHWT224	HWT93	HWT97L	HWT97X	Household weight based on 7,224 HHs interviewed in IFLS1 and all HHs interviewed in IFLS2.
HHWT730	HWT93SMP	—	—	Household weight based on 7,730 HHs listed in IFLS1 target sample. There is no corresponding weight in IFLS2.
ROSTERWT	PWT93	PWT97L	PWT97X	Person weight based on all individuals listed in a HH roster, adjusted for HH selection probabilities. In IFLS2, all individuals were supposed to get individual books so this weight also applies to individual book respondents.
RESPWT	PWT93IN	PWT97INL	—	Person weight for the IFLS1 "Main" respondents who were administered an individual book. Use these weights when using responses from individual books (B3, B4 and B5) in IFLS1 or when using IFLS1 and IFLS2 in combination and using only the "Main" respondents. There is no corresponding cross-section weight.
CA_WT	PWT93US	PWT97USL	PWT97USX	Person weight for anthropometry and health assessments in IFLS1 and IFLS2.
—	—	—	PWT97EKX	Person weight used for cognitive assessments in IFLS2.

All weight variables are stored in HTRACK (for HH level weights) and PTRACK (for individual level weights).

Longitudinal analysis weights adjust baseline weights for attrition. Statistics that are weighted with these variables should reflect the 1993 distribution of individuals and households in the 13 IFLS provinces.

Cross-section analysis weights take into account attrition and changes in the population distribution between IFLS1 and IFLS2. They are intended to reflect the distribution of individuals and households in the 13 IFLS provinces in Indonesia at the time of IFLS2.

Table 3.4
Probability a HH is recontacted in IFLS2: Logit estimates

	β	[se]
$\ln(\text{per capita expenditure})$ spline		
-- 1st quartile	0.330	[0.207]
-- 2nd quartile	-0.654	[0.441]
-- 3rd quartile	-0.080	[0.354]
-- 4th quartile	-0.280	[0.131]
HH size	0.119	[0.037]
(1) if 1 person HH	-0.986	[0.208]
(1) if 2 person HH	-0.465	[0.188]
Location in 1993		
(1) if urban	-1.043	[0.133]
(1) if North Sumatra	-0.419	[0.189]
(1) if West Sumatra	0.035	[0.262]
(1) if South Sumatra	-0.304	[0.232]
(1) if Lampung	-0.176	[0.309]
(1) if West Java	0.722	[0.198]
(1) if Central Java	1.959	[0.346]
(1) if Yogyakarta	0.870	[0.241]
(1) if East Java	0.686	[0.212]
(1) if Bali	0.254	[0.278]
(1) if West Nusa Tenggara	1.654	[0.474]
(1) if South Kalimantan	-0.281	[0.246]
(1) if South Sulawesi	0.355	[0.292]
Intercept	1.978	[0.685]
Pseudo R ²	0.119	
Sample size	7,224	

Notes: Sample is all HHs interviewed in IFLS1. All covariates are measured in 1993.

Table 3.5 Probability an individual is recontacted in IFLS2: Logit estimates

	Target respondents		Other respondents	
	β	[se]	β	[se]
<i>Respondent characteristics</i>				
(1) if head of HH in 1993	1.104	[0.117]	.	
(1) if spouse of head of HH in 1993	1.397	[0.121]	.	
(1) if main respondent in 1993	0.218	[0.110]	.	
(1) if child of main respondent in 1993	.		0.378	[0.126]
(1) if child of head of HH in 1993	0.678	[0.085]	0.875	[0.062]
Age in 1993 (spline)				
-- 0-10 yrs	-0.012	[0.016]	-0.042	[0.017]
-- 10-15 yrs	-0.337	[0.031]	-0.310	[0.030]
-- 15-20 yrs	0.107	[0.035]	-0.077	[0.018]
-- 20-30 yrs	0.038	[0.015]	0.059	[0.022]
-- 30-45 yrs	0.045	[0.009]	0.264	[0.176]
-- 45-60 yrs	0.026	[0.011]	-0.272	[0.217]
-- >60 yrs	-0.005	[0.011]	0.436	[0.528]
<i>Household characteristics</i>				
(1) if 1 person HH	-0.835	[0.170]	.	
(1) if 2 person HH	-0.430	[0.114]	0.061	[0.234]
# HH mems age 0-9	-0.001	[0.027]	-0.069	[0.024]
# HH mems age 10-14	0.145	[0.035]	-0.014	[0.029]
# HH mems age 15-24	0.136	[0.026]	-0.015	[0.017]
# HH mems age >=25	0.120	[0.032]	0.173	[0.026]
Years of education of head	-0.034	[0.008]	-0.027	[0.008]
Years of education of spouse	-0.028	[0.009]	-0.028	[0.009]
(1) if spouse exists	0.174	[0.080]	-0.041	[0.072]
□n(PCE) spline				
-- up to 3rd quartile	0.087	[0.051]	0.004	[0.050]
-- top quartile	-0.170	[0.067]	-0.160	[0.077]
<i>Survey characteristics</i>				
# HHs in EA interviewed in 1993	-0.063	[0.041]	-0.055	[0.033]
% target HHs in EA completed in 1993	0.030	[0.009]	0.017	[0.008]
1993 interviewer assessment				
(1) if HH provided excellent answers	0.197	[0.089]	.	
(1) if HH provided good answers	0.188	[0.057]	.	
<i>Location in 1993</i>				
(1) if urban	-0.902	[0.400]	-0.357	[0.323]
(1) if North Sumatra	-0.169	[0.093]	-0.784	[0.109]
(1) if West Sumatra	0.583	[0.134]	-0.245	[0.127]
(1) if South Sumatra	0.287	[0.119]	-0.520	[0.127]
(1) if Lampung	0.271	[0.141]	-0.646	[0.149]
(1) if West Java	1.080	[0.100]	0.165	[0.102]
(1) if Central Java	1.062	[0.112]	-0.304	[0.108]
(1) if Yogyakarta	1.317	[0.150]	-0.443	[0.127]
(1) if East Java	0.734	[0.098]	-0.134	[0.111]
(1) if Bali	0.625	[0.136]	-0.499	[0.145]
(1) if West Nusa Tenggara	0.987	[0.148]	-0.449	[0.131]
(1) if South Kalimantan	0.299	[0.131]	-0.188	[0.144]
(1) if South Sulawesi	0.456	[0.130]	-0.389	[0.123]
Intercept	-0.154	[0.486]	1.794	[0.425]
Pseudo R ²	0.112		0.150	
Sample size	23,948		9,133	

4. Special Features of the IFLS Data

This section discusses the distinctive features of IFLS2 data as they affect analysis files. The bulk of the discussion applies to the HHS, with the CFS covered at the end of the section.

Symmetric Information

In two IFLS2 modules, HR (assets) and PK (decision-making), husbands and wives provided symmetric information. That is, a husband answered questions about himself and about his wife, and the wife answered the same questions about herself and about her husband. These data allow comparisons of partners' perspectives about themselves and their spouses.

In module KW, individuals provided information about the dates of their marriages and gifts given and received at the time of marriage. Within a household, if two individuals are married to each other, their KW data could be compared for consistency. Or, if one individual's data is missing, data from the spouse could be used to fill the gap. Similarly, in module MG, individuals described their migration experiences. If couples or parent-child pairs had moved together and each individual answered MG, their responses could be compared to check consistency, or the MG data of one could supply information missing from the other.

Duplicate Information

Certain pieces of information were collected in more than one place. In most cases, the respondent was one source of information and a proxy respondent (or preprinted information) was the other. For example, the household roster (module AR) contained information on a number of topics that were also in the questionnaire books addressed to individuals. Though it would be easier to use the information from the roster, data from the individual books are likely to be more accurate, since the information was self-reported rather than provided by proxy.

Age. Information on age was collected in both the AR roster (generally by proxy) and on the covers of the individual books. In addition, in certain places in the questionnaire, interviewers were required to examine the age recorded on the book cover, usually to determine whether the respondent was above a certain threshold age. We did not correct inconsistencies between the roster and book covers, but the PTRACK file contains a "best-guess age" variable. We did not attempt to correct inconsistencies between the roster and questions within the book, since complicated skip patterns were often involved.

Birthdate. Information on birthdate was collected in individual books and by the nurse who conducted the health assessment (book US). For new respondents, birthdate was also recorded in the AR roster. We did not correct inconsistencies between the AR roster and the health assessment, but the PTRACK file contains a best-guess birthdate variable. If the respondent knew the year of birth but not the month or day, this variable shows month and day as 98.

Sex. A respondent's sex was preprinted in the AR roster and collected on the cover of books 3A and 3B. In cases of inconsistency between the roster and book covers, we undertook extensive checks on name, IFLS1 information, and other data to ascertain a best guess for sex. The best-guess sex values are recorded in PTRACK, in the AR roster (AR07), and on the individual book covers.

Marital Status. Marital status was noted in the AR roster and on the covers of books 3A, 3B, and 4. Various interviewer checks within the individual books required using marital status information from

the book cover. In cleaning the data, we tried to make sure that marital status in the roster matched marital status on the book covers. We did not clean interviewer checks because that would have required complicated adjustments to skip patterns. 33

Education Level. The AR roster reported the highest level of schooling attained and the highest class completed within that level (AR16 and AR17). For many respondents that information is repeated in book 3A or 5.

Earnings and Nonlabor Income. Module TK asked in depth about employment and labor earnings. The proxy book also addressed these topics. As insurance in case neither module was completed for some household members, we also included a question on earnings in the AR roster (AR15b). The existence of the AR data means that a measure of total household labor income can be computed, even if not all household members provided a book 3A or proxy book. However, data from TK (or book Proxy) are preferred because they come from the respondent or a knowledgeable proxy. TK data are likely to be more accurate also because earnings were addressed in the context of related questions.

Book 2, module HI, asked about nonlabor income at the household level, and book 3A, module HI, asked about it at the individual level. The individual-level HI information is preferred, but the household summary is useful for computing total household income if an individual book is not available for all adults.

Parents' Survival Status. The AR roster recorded PID numbers for each individual's mother and father (AR11 and AR10). If the mother or father was not a member of the household, codes were used to designate whether the parent was alive and living in another household or dead. Book 3B, module BA (parent) explicitly asked the respondent about each parent's survival status. The BA data are preferred.

Timing of Marriage. Both the KW and KL calendars provide data on the timing of a woman's marriages.

Timing of Pregnancy. Both the CH module and the KL calendar provide data on the timing of a woman's pregnancies.

Current Method of Contraception. Both modules KL and CX provide information on whether a couple is currently using a method of contraception and if so, what method is used.

Family Relationships

The IFLS contains extensive information on family relationships, particularly between husbands and wives and between parents and children. The information is not limited to household members but also covers non-coresident kin.

Parents, Children, and Spouses Identified in the AR Roster

The AR roster provides much information on relationships among current household members, as shown in the table below:

Variable	Information	Remarks
AR02	Which member was designated household head in IFLS1 and how other members in IFLS1 were related to that person	Sometimes this information indicates how members other than the household head were related. For example, if persons 3 and 4 were both children of the head, they were either full or half-siblings. If person 4 was the mother of the head and person 3 was the child of the head, person 4 was almost certainly the grandmother of person 3. In other cases the information is not definitive. For example, if persons 5 and 6 were both grandchildren of the head, they were likely to be siblings or cousins, but we do not which from AR02/02b.
AR02b	Which member was designated household head in IFLS2 and how other members in IFLS2 were related to that person	
AR10, AR11, and AR14	PID numbers of an individual's birth father, birth mother, and spouse	To find the education level of a child's parents, use the line numbers in AR10 and AR11 to link child to parents and thus to parents' education data either in the AR roster or in their individual books 3A. If a person's mother, father, or spouse was alive but not a household member, AR10/11/14 = 51. If a person's mother or father was dead, AR10/11 = 52. If a person's spouse was dead, the person was a widow and skipped AR14.

Note two cautions in using the AR data on family relationships. First, because the household rosters were preprinted, a person's father/mother/spouse sometimes has a line number in the roster (indicating that they lived in the household in 1993) but was not a current member of the household (had moved or died before the 1997 interview). If so, AR10/11 = 51. Second, the accuracy of codes 51/52 is not clear. The person completing the roster may have known that the father or mother was not in the household but not whether the father or mother was living or dead. For parents' survival status, book 3B, module BA (parent), is the preferred source of information because an explicit question was posed directly to the respondent.

Parents, Children, and Spouses Identified in Other Modules

Information about parents, children, and spouses in modules other than AR is described in the table below. That information usually applies to relatives who were not current members of the household (that is, they are non-coresident kin). If a relative was a current member of the household, his or her PID and other characteristics were on the AR roster, and he or she probably filled out an individual-level book.

Module	PID and Other Identifiers	Other Information
KW (marital history)	PID of current or most recent spouse, if a current household member	Education level of current and previous spouses if not current household members

Module	PID and Other Identifiers	Other Information
BA (parents)	PID of mother and father if current household members	Age, education, marital status, occupation of non-coresident parents
BA (preprinted sibling roster*) *Rosters that were not preprinted address only non-coresident siblings, who do not have PID numbers	PID not included, some information on line numbers in IFLS1 sibling roster (see Appendix C, module BA)	Age, education, marital status, occupation, location of residence
BA (preprinted child roster*) *Rosters that were not preprinted address only non-coresident children, who do not have PID numbers	AR roster number in IFLS1, CH (pregnancy) number in IFLS1, BA line number in IFLS1	Age, education, marital status, occupation, location of residence
CH (pregnancy history)	PID if current household member	Details of each pregnancy and birth, breastfeeding, survival status Age, education, marital status, occupation, location of residence for some children

Non-coresident Siblings. Module BA (sibling) in book 3B provides the most detailed information about non-coresident siblings. It was not necessary to collect information on siblings' characteristics from all respondents in the household if the information had already been provided by another respondent. For example, if person X was a sibling of the household head, he or she had the same siblings as the head. In that case, information provided by the head about his or her siblings serves as sibling information for person X as well. If person X was a child of the household head, person X's siblings were the other children of the head. Since the household head reported about his or her children, that information can serve double duty for the head's siblings and children:

Relationship to Household Head or Spouse	Location of Information on Non-coresident Siblings
Brother or sister of head	Sibling roster for head
Brother or sister of head's wife	Sibling roster for spouse of head
Child of head	Child roster for head
Child of head's wife	Child roster for head's wife
Other	Own sibling roster

Non-coresident Children. Module BA (child), whether in book 3B or 4, provides the most detailed information about non-coresident children. It was not necessary to collect information on children's characteristics from all respondents in the household if the information had already been provided by another respondent. For example, if person X (a man) was married to person Y and had had no other wives, his children were also the children of person Y and only person Y was asked to report about them. If person X had additional children with another wife, however, he was asked to report about those children himself.

Classifying Relatives

Some relationships were not always specified with precision. In particular, the distinction between biological and through-marriage relationships was sometimes blurred. It was not always clear whether a child/parent was a biological child/parent, a step-child/-parent, or a child-/parent-in-law. Nor was it always clear whether someone classified as an aunt or cousin was related to the respondent or the respondent's spouse. We did not attempt to resolve all such inconsistencies. They were likely to arise in the contexts described below.

AR02b vs. AR10/11. Occasionally AR02b classified someone as the child of the head, but AR10 or AR11 did not list the head as the person's biological parent. The reason may be that AR10/11 asked specifically about the biological parent, whereas AR02b asked more generally about the relationship to the head. Likewise, AR02b sometimes listed an individual as the parent of the household head, but that person's PID did not appear in the head's response for AR10/11 as a biological parent of the head.

Divorce between Survey Waves. Between IFLS1 and IFLS2 some marriages ended in divorce. During the "other" cleaning process (see Sec. 5), we found responses indicating that someone was an ex-spouse or related to an ex-spouse. We created two new categories (ex-spouse and relative of ex-spouse) to account for this.

Asset Ownership. Modules HR, UT, and NT contained questions asking whether other family or household members were co-owners of various assets. In some cases it is not clear whether someone categorized as an aunt is related to the respondent or the respondent's spouse.

Identifying All of a Person's Closest Relatives

To count the total number of children, siblings, or living parents for a respondent, or to obtain information on the characteristics of these kin, it is necessary to merge information from several modules and sometimes to draw information from IFLS1 data. Table 4.1 provides some pointers.

CFS: Using Information from Multiple Respondents

Within the CFS, several types of data are available from multiple "respondents" per community: data on prices, data from health facilities and schools, and data from community informants on the availability of services and on sanitation and infrastructure. For some analyses it will be useful to combine the data from multiple respondents to reduce measurement error of certain constructs or to produce an aggregate value for the community as a whole.

For example, both the head of the community and the head of the women's group answered modules I and J on the availability of schools, health facilities, and health outreach programs. If data for a particular question are missing from one respondent, data from the other respondent can be used to supply the missing information.

Data on prices of food and nonfood items are available from the respondents to book PKK, respondents to book Posyandu, and from visits to markets and sales outlets recorded in book 2. Within a community, uniquely identified by variable COMMID97, data on prices are available from up to six community-

level informants. Users may wish to construct prices for the community by calculating mean or median³⁷ prices across these six informants.

Data from facilities are available for 2–4 facilities per community. For some analyses it may be useful to construct measures of average (or median) service prices or quality at the facilities that serve a particular community. To make these calculations, the analyst will need to determine which facilities are available to the community. This information is provided in the SAR. For each community, the SAR contains a list of facilities mentioned by household and community-level informants as service options in that community. Each facility mentioned in the SAR has an FCODE. By merging the data from the SAR with the data in the facility files (merge on FCODE), information from the facility files will be added to the list of available facilities in the SAR. One can then compute the average characteristics of the facilities associated with each community.

Note two caveats. First, not all facilities identified in the SAR were interviewed (as discussed in the *Study Design*, DRU-2238/1-NIA/NICHD, a quota of facilities was interviewed in each community). Therefore, a number of observations in the SAR will not match to any of the facility data. Second, certain facilities appear more than once in the SAR because the facilities were available to more than one IFLS community. Within the SAR, the combination of COMMID97 and FCODE uniquely defines an observation. FCODE alone does not uniquely define an observation.

Sources of Information for a Respondent's Closest Relatives

Siblings	
In the household	<ul style="list-style-type: none"> • For a household head, use AR02b to identify household members who are brothers and sisters. • For a non-head whose mother or father is a household member, check the roster for other individuals in the household who identify the same parents in AR10/AR11. • For a non-head whose parents are not household members, sibling information is unclear.
Outside the household	Use the respondent's BA (sibling) data from book 3B
Children	
Of new female respondents	For children ever born, use BR data.
Of panel female respondents	For children ever born, use IFLS1 BR data if the woman was older than 49 in IFLS1. Or combine IFLS1 BR data with IFLS2 CH data on pregnancies after 1993 if the woman was 49 or younger in IFLS1.
Of male respondents	For children with current wife, use information provided by wife. For non-coresident children older than 15 born to a previous wife, use BA (child) data from book 3B. In IFLS2 there is no information about non-coresident children younger than 15 born to someone other than the current wife. For panel respondents, this information should be available in IFLS1.
Parents	
In the household	Use data from books 3A and, 3B completed by each parent (preferred) or from the respondent's AR information.
Outside the household	Use the respondent's BA (parent) data from book 3B.

5. Cleaning the IFLS Data

This section describes the procedures carried out during the fielding period and afterward to minimize errors in the IFLS2 data. Additional information on survey operations is provided in the *Study Design* (DRU-2238/1-NIA/NICHD), Appendix A.

In the Field: CAFÉ Editing, Interviewer Rechecks

Data cleaning began in the field. Interviewers filled out the paper questionnaires while in the respondents' households, then edited their work at base camp. For both the HHS and CFS, interviewers were responsible for turning in legible questionnaires that had been filled out as completely and accurately as possible.

A process of Computer-Assisted Field Editing (CAFÉ) was used to help maintain data quality in the HHS data.¹⁸ Interviewers handed in their completed paper questionnaires to a CAFÉ team at base camp. The CAFÉ team entered and edited the data on laptop computers, using data-entry software designed to detect a variety of fielding errors. Range checks identified illogical values, such as a sex value of 2 when sex was supposed to equal 1 or 3. Cross-book checks identified more complex inconsistencies. For example, if the sex listed for a respondent in the AR household roster was inconsistent with the respondent's sex recorded on the cover of book 3A, an error message was generated.

The CAFÉ editor was responsible for resolving error messages with the interviewer. Some errors could be resolved fairly easily. For example, the interviewer might remember the sex of a respondent interviewed earlier in the day and verify that the inconsistency was due to a careless error. Other errors required the interviewer to return to the household and check with the respondent. For example, if the screening questions at the beginning of module RJ (outpatient visits) recorded that the respondent had used both public and private services but the detailed questions recorded only a public-sector visit, the interviewer might need to go back to the household to determine whether a visit to a private provider had occurred and if so, collect more information on the visit.

The CAFÉ team was critical to the collection of high-quality data in IFLS2. When its work was finished for an enumeration area, the data were sent to the Jakarta office and were electronically transmitted (via ftp) to RAND in Santa Monica. A team there performed basic data quality checks, monitored recontact rates, and provided feedback to the teams in the field.

In Jakarta

Double Data Entry and Verification

We followed a standard procedure for eliminating transcription errors by entering the data from the paper questionnaires twice and then comparing the two sets of data. For HHS data, the work of the CAFÉ teams served as the first entry; the second entry was done in the Jakarta office. For CFS data, both first and second entries were done in Jakarta. The two electronic versions of the data were compared and all discrepancies manually verified against the paper questionnaire. If an error occurred in the version of the data set that was to serve as final, the data were corrected.

¹⁸ Resources were not sufficient to use CAFÉ for the Community-Facility Survey.

For detecting and resolving more complicated errors, we implemented a “Look Ups” (LU) cleaning process. It involved the use of a sophisticated, customized computer program to run checks, with followup of suspected errors by specialists with extensive field experience, who consulted the paper questionnaires. The LU phase was important to quality assurance because

The paper questionnaires sometimes contained valuable written information that was not captured in the electronic data. For example, an inconsistency might be generated because an editor had made an inappropriate correction. Reference to the interviewer’s original annotation resolved the issue so the data could be corrected.

LU specialists were drawn from our best interviewers, editors, and field supervisors. We wanted to capitalize on the expertise they had gained in fielding the survey to help resolve more difficult issues before releasing the data for analysis.

As the questionnaires contained many related questions, it was sometimes easier and faster to check responses in the questionnaires themselves than to program computerized checks.

The LU program ran checks within and across questionnaire books for a particular household.¹⁹ Some checks repeated CAFÉ procedures, in an effort to resolve inconsistencies that remained after CAFÉ editing. More complicated checks were added as a result of discoveries made when the data were first checked in Santa Monica. For example, a number of pregnancies were recorded to have produced a multiple birth, but there was no evidence that twins had actually been born. An LU check was set to generate an error message if a pregnancy was recorded as resulting in a multiple birth on a certain date but no other births were shown for that date. For each case flagged, the LU specialist then examined all related data (AR roster, BR pregnancy summary, CH pregnancy history) to determine whether the multiple birth report was accurate.

To give other examples, the LU program also checked that

no individual-level books were filled out for IFLS1 household members reported as dead or departed in the IFLS2 AR roster (AR01a). If such a book existed, the specialist had to ascertain whether AR01a was incorrect or the PID on the individual book was incorrect.

the last place to which a respondent reported moving was the respondent’s current residence.

parents were at least 12 years older than their children.

For each error message generated, the LU specialist was required to check the problem on the paper questionnaires and record in a log file whether and how the problem could be corrected and whether a correction was in fact made. If the specialist was not sure how to correct the data, the data were not to be changed but a suggestion could be entered in the log file. Some problems were relatively straightforward to correct. Others, such as skip patterns that weren’t followed, could not be corrected because the data had not been collected.

In training and supervising the LU specialists, we repeatedly stressed that specialists could not make up data, change an answer simply to force consistency, or correct errors they believed the respondent had made. Instead, specialists were to look for evidence of the correct answer on the paper questionnaires where an interviewer or data entry error was suspected. As a result, not all inconsistencies were corrected during LU; many were addressed later in Santa Monica.

¹⁹ Look Ups checks per se were not run for CFS data as it would have been impractical to write a Look Ups program given the much smaller number of observations in the CFS. The Notes part of the LU work, described below, did cover CFS data.

In various places throughout the HHS and CFS questionnaires, interviewers were asked to comment if they believed a response warranted explanation, clarification, or correction. We judged it important to capture any suggestions in these notes for correcting the data. Accordingly, for both HHS and CFS data, we trained two “Notes” teams of specialists to generate an electronic file of suggested corrections to the data from interviewers’ notes (including the CP modules at the end of nearly every HHS book. For HHS data, the suggestions were reviewed by the LU specialists and carried out if the specialist agreed. For CFS data, a Notes team implemented necessary changes. 41

Both Look Ups and Notes staff received extensive training and supervision to ensure an extremely conservative approach to changing the data and to ensure the proper recording of all changes (and suggested changes) so that they could be reviewed and undone later if necessary.

Special Cleaning for Open-Ended, “Other,” and Numeric Variables

Open-Ended Variables. The questionnaire elicited open-ended responses for some questions that did not lend themselves to closed-ended responses. Cleaning for those responses involved providing rough English translations (minus any information that might be used to identify the respondents). Knowledge of *Bahasa Indonesia* was required, and the cleaning was done by one of two specially trained teams in Jakarta.

Variables with “Other” Answers. “Other” answers occurred when a response varied from the pre-coded options. In cleaning “other” responses, it was necessary to review the text responses and decide whether a response could be coded into an existing category, whether creation of new category was warranted, or whether the response should remain coded as “other.” Knowledge of Bahasa Indonesia was required, and the cleaning was undertaken by one of two specially trained teams in Jakarta.

New categories were typically created if a response was substantively different from the pre-coded responses and it occurred a non-trivial number of times. When new categories were created, they were assigned a code larger than the existing “other” code, indicating that the category had not existed as a response option in the fielded questionnaire. We were inclined to create new categories rather than leave a large “other” category. Users thus have the option of aggregating the data, whereas finer *disaggregation* of the data would be impossible if new codes were not created.

Three types of “other” variables were cleaned:

Simple questions allowing only one answer (e.g., highest education level completed). “Other” responses were recoded to a new or existing response category.

Questions where multiple responses were allowed (such as which family members were co-owners of a particular asset). “Other” responses were recoded to a new or existing category, and the indicator that an “other” response had originally been selected was turned off. For example, suppose that for question NT04 a respondent reported that he co-owned a business with his sibling, but he also mentioned in the “other” category that his sister was a co-owner. The original answer would have taken the value FH (sibling + other). The recoded answer would just take the value F because the response categories do not distinguish brothers from sisters but group them as siblings.

Questions that related to items in a grid. Cleaning of “Other” responses here might generate another item in the grid. For example, module PS asked about self-treatment with various medicines. A number of “other” respondents reported having used vitamins, so “vitamins” was added to the grid.

When new items were added to the grid, the value for the question asking respondents whether there was an “other” item was reset to No, and that value for the new response category was set to Yes. For respondents who answered that there was “no” other item, the value for the new response categories was typically set to 4, indicating that the respondent was not directly

asked this question. Continuing the PS example, the table below illustrates the range of changes that were involved when new items were added to a grid:

42

	Before Cleaning	After cleaning	Explanation
PSTYPE	A B C D Other	A B C D Other E Vitamins F Refreshers	Two new categories (Vitamins and Refreshers) were added.
No. of cases where PS01 = Yes, by PSTYPE	A: 10,425 B: 4,001 C: 4,294 D: 240	A: 10,462 B: 4,052 C: 4,308 D: 8 E: 7 F: 52	Frequencies changed because some "other" answers were recoded to an existing category or to a new category.
Response codes: PS01 when PSTYPE = E or F	PSTYPE never = E or F	PS01 = 1 Yes if "other" answer recoded to E or F PS01 = 4 Not asked	Respondents were not explicitly asked whether they self-treated with vitamins or refreshers.

Numeric Variables. Some numeric responses did not fit the space provided, either because the answer had too many digits or required more decimal places than were allowed. In these cases, interviewers had been trained to fill the space provided with a string of 9's ending in a 5 ("out of range") and to record the correct answer in the "Notes" section of the questionnaire or in the "other" answers file. If warranted by the interviewer's annotations, we widened the numeric field to allow the correct answer and replaced the "out of range" code with the correct answer. It was not possible to correct all out of range codes, so *special codes* sometimes still appear in the data.

In Santa Monica

In Santa Monica we did additional cleaning to correct remaining errors and to make the publicly available files as easy to use as possible.

Module Checks

For each data module, we made an effort to

- Review the LU checks and determine whether any remaining errors or inconsistencies could be corrected.

- Review numeric responses for the existence of special codes and review character variables for responses meaning "empty" or "don't know" in Bahasa Indonesia.

- Create or correct X variables (defined in Sec. 3) so that the special codes were preserved and the associated numeric or character variable contained only valid responses.

Check that skip patterns were properly followed and apply corrections if data would not be lost as a result.²⁰ 43

Check that TYPE variables (defined in Sec. 3) exist in grids.

Assign variable names and labels as clearly as possible.

Check for and document any cases that stood out as particularly odd or unusual.

Find and drop any variables that might help identify a respondent.

Checks on IDs across Books and Survey Waves

It is essential that IDs such as HHID97, PID97, and PIDLINK (defined in Sec. 6) be correctly assigned. Therefore, we rigorously checked ID assignments. For example, when two very different ages were reported for the same individual (e.g., in the AR roster and on an individual book cover), the case was reviewed to determine whether PID97 had been correctly assigned in each place. Similarly, correct PIDLINKs for members of split-off households are necessary in order to identify whether or not each member was also a member of a 1993 household. Three people independently reviewed the assignment of PIDLINKs in split-off households, and all inconsistencies in their reviews were reconciled.

Checks on Book Covers

A number of checks were run to verify that the information on book covers was as accurate as possible. For example, for the Proxy Book and book 5, where someone other than the respondent was likely to have answered, we checked to make sure that the relationship of the actual respondent to the intended respondent matched information in the AR roster on the relationship between those individuals. For books K, 1, and 2, if a PID was given that corresponded to someone younger than 18, we checked the name to make sure that the PID was recorded correctly. In some cases a child or young adult had provided information in books K, 1, or 2; in other cases PID had been recorded incorrectly. We also ran additional consistency checks for items like sex and age that appeared in both the AR roster, on the individual book covers, and within the individual books, and made corrections where the weight of the evidence indicated the correct answer.

Checks on Preprinted Child and Sibling Rosters

We checked whether the preprinted rosters that existed were used and created flags indicating cases where a roster existed but was not used. When a preprinted roster existed but was not used, the analyst should not match IFLS1 BA data to IFLS2 BA data solely by the line number of the child or sibling, since the listing order was probably different in the two waves.

Checks on Units of Measure

Some questions asked for a numeric answer and allowed a choice of units of measure. For example, PP5 asked the travel time to health facilities, allowing the answer in minutes, hours, or days. Occasionally respondents provided answers that were clearly outliers. They were reviewed with other information, such as the location of the facility, to ascertain the correct unit. For example, if a respondent said that it took 10 hours to get to a traditional practitioner but we found the practitioner to be located in the same village as the respondent, we judged the proper unit to be minutes rather than hours. Similarly, if a

²⁰ The IFLS2 questionnaires contained a number of complicated skip patterns that controlled the flow of the interview. Interviewers did not always follow these patterns correctly, so for some modules, some respondents provided either more or less information than was necessary. Generally we did not correct skip patterns, since we did not want to delete information (even if it was collected in error), and there was no way of generating a response when the question had not been asked.

woman reported a miscarriage after a pregnancy of 11 months, we judged the proper unit to be weeks rather than months. Such corrections typically involved very few cases, usually fewer than 25. 44

Created Variables and Files

We created some variables and files to make the data easier to use. For example:

Variable MOVE summarizes the information on a household's current location relative to its 1993 location (or, for split-off households, the origin household's 1993 location).

Files HTRACK AND PTRACK indicate what data are available for households and individuals (respectively) in each survey wave.

Variable PPCHILD indicates whether a PP child roster was used. If so (PPCHILD = 1), a line number in the IFLS2 child roster refers to the same individual listed for that line number in the IFLS1 child roster.

6. Using IFLS2 Data with IFLS1 Data

This section provides guidelines for using both waves of IFLS data to obtain longitudinal information for households, individuals, and facilities.

IFLS1 Re-Release

A revised version of IFLS1 has been prepared to facilitate use of the IFLS1 and IFLS2 data together. Abbreviated IFLS1-RR (1999), the re-release incorporates the following major revisions:

Adjustments outlined in the “fixes” files have been incorporated.

Subfiles with the same unit of observation have been joined, as is desired by most users.

IFLS2 identifiers (HHID93, PIDLINK, COMMID93, and FCODE—explained below) have been added to facilitate linking the IFLS1 and IFLS2 data.

The restructured data are designed so that the existing IFLS1 codebooks can still be used. No variable names have been changed; a few new variables have been added. File names are the same, except that some no longer appear because subfiles have been combined into a new file. In general, the name for the new combined file reflects the name of the first subfile in the series of files that were joined. For example, the new file named BUK3TF1 is a combination of former subfiles named BUK3TF1 and BUK3TF2.

The documentation for IFLS1-RR describes the fixes applied to the IFLS1 data, the new variables added, and the new files created by merging related IFLS1 subfiles.²¹ For quick reference, tables list all the IFLS1-RR subfiles and their contents. Because IFLS2 uses a slightly different file naming convention, and modules were added or dropped compared with IFLS1, a set of tables shows how IFLS1 subfiles map to their IFLS2 counterparts.

Differing IFLS1 and IFLS2 Household IDs

In IFLS1 the household ID was called CASE and was 9 digits long, with groups of digits assigned the following meanings:

<u> x x </u>	<u> x x </u>	<u> x x x </u>	<u> x x </u>
province	<i>kabupaten</i>	EA	specific household

Since the EA already defines the province and *kabupaten*, the first four digits are superfluous. In IFLS2, we assigned each household a 7-digit ID called HHID97, with the digits carrying the following meanings:

<u> x x x </u>	<u> x x </u>	<u> x x </u>
EA	specific household	origin/split-off

In the last two digits, 00 designates an origin household. For a split-off household, the 6th digit is always 1, signifying a split in IFLS2, and the 7th digit indicates whether it is the first, second, or other split-off (some multiple split-offs occurred).

²¹ Christine E. Peterson, *Documentation for IFLS1-RR: Revised and Restructured Indonesia Family Life Survey, Wave 1*, RAND, DRU-1195/7-NICHD.

For example, consider hypothetical household 327412501 in IFLS1. By IFLS2, the head had divorced and moved out, and one of his sons had married and moved out. We interviewed the divorced wife in the original home, the divorced head in his new household, and the son in his new household. The resulting IDs for IFLS2 would be as follows:

HHID97	Target Respondent	Type of Household
1250100	Wife	Origin household
1250111	Head or son (whichever we found first)	First split-off household
1250112	Head or son (whichever we found first)	Second split-off household

Merging IFLS1 and IFLS2 Data for Households and Individuals

The method for merging household-level information depends on whether the original or re-released version of the IFLS1 files is used:

<i>Original IFLS1</i>	<i>IFLS1-RR</i>
Create an HHID93 for each IFLS1 household that matches HHID97 in IFLS2. To do so, drop the first four digits of CASE and add 00 at the end (SAS and Stata code shown below). For example,	Rename HHID93 "HHID97." Match IFLS1 and IFLS2 data on HHID97.
<div style="text-align: center;"> CASE 327412501 </div> → <div style="text-align: center;"> HHID93 1250100 </div>	

Not all households will merge. Some IFLS1 households were not reinterviewed in IFLS2. And households that were new in IFLS2 will not have data in IFLS1.

To merge individual-level information, note that the person identifier in IFLS2, called PID97, corresponds to PERSON in IFLS1 and PID93 in IFLS1-RR. To merge a person's information from two different IFLS2 books, merge on both HHID97 and PID97. To merge a person's information from two different IFLS1-RR books, merge on both HHID93 and PID93.

HHIDxx and PIDxx are *not* used to link individual information between IFLS1 and IFLS2.

The variable PIDLINK is essential for merging individual-level data between IFLS waves for persons who were in both 1993 and 1997 households. PIDLINK is a 9-digit identifier consisting of the following:

x	x	x	x	x	0	0	x	x
1993 EA			1993 household		origin		PERSON [1993]	

For analysts using the original IFLS1 files, the following SAS or Stata code may be used to obtain the household and person identifiers needed for merging the two waves of data: 47

SAS Code:

```
pi d93 =person;
hhi d93 =compress(substr(put(case, z9. ), 5, 5) || "00");
pi dl i nk=compress(substr(put(case, z9. ), 5, 5) || "00" || put(person, z2. ));
```

Stata Code:

```
gen pi d93=person
gen str7 hhi d93=string((mod(case, 100000) * 100), "%07.0f")
gen str9 pi dl i nk=hhi d93+string(pi d93, "%02.0f")
```

PIDLINK has nothing to with the household in which the person was found in 1997. Continuing the above example of the divorced household, suppose that in IFLS1 the head's PERSON number was 01, his wife's number was 02, and their son's number was 03. Assume that in IFLS2 the husband was contacted before the son. The range of identifiers for these individuals would be as follows:

	CASE	PERSON	HHID93	PIDLINK	HHID97	PID97
Husband	327412501	01	1250100	125010001	1250111	01 (in split-off household)
Wife	327412501	02	1250100	125010002	1250100	02 (same as 93—still in origin)
Son	327412501	03	1250100	125010003	1250112	01 (in split-off household)

This example illustrates another point. Some PIDLINKs appear in two different IFLS2 household rosters: the preprinted roster of the origin household, in which the individual has AR01a = 3 (moved out of household), and in the household roster of the split-off household to which the individual was tracked and interviewed. To avoid duplicate PIDLINKs when merging data from the AR roster with IFLS1, drop AR records where AR01a = 3.

Data Availability for Households and Individuals: HTRACK and PTRACK

Files named HTRACK and PTRACK indicate what data are available for households and respondents, respectively, in each survey wave.

HTRACK

HTRACK contains a record for every household that was interviewed in IFLS1 or IFLS2. There are 8,116 household-level records in HTRACK, one record for each of the 7,224 households that were interviewed in IFLS1 and one record for each of the additional 892 split-off households that were added in IFLS2. HTRACK provides information on whether the household was interviewed in either wave and, if so, whether data from books K, 1, 2, and US are available. HTRACK also provides information on the household's location in 1993 and in 1997. For 1993, two sets of location codes are given: those used by the Central Bureau of Statistics (BPS) in 1993 (also in the original IFLS1 data), and those used by BPS in

1998.²² For 1997, only the codes in use as of 1998 are given. The codes in use by BPS in 1998 are used consistently throughout IFLS2. 48

For households that were interviewed in 1997, variable MOVER97 identifies whether the household moved between 1993 and 1997, taking the following values:

- 0 = Did not move
- 1 = Moved within same village/municipality
- 2 = Moved within same *kecamatan*
- 3 = Moved within same *kabupaten*
- 4 = Moved within same province
- 5 = Moved within other IFLS province

MOVER97 is non-missing not only for origin households interviewed in 1997 but also for split-off households. Because each split-off household contains at least one person who was tracked from an origin household in 1993, we have calculated MOVER97 for split-off households on the basis of the household's 1997 location relative to the location of the origin household in 1993, from whence the tracked individual came. Similarly, the variables for the data on location in 1993 are non-missing for split-off households, even though these households were not interviewed in 1993. For split-off households, the 1993 location information reflects the location of the origin household that generated the split-off household.

In addition to the BPS location codes, HTRACK contains COMMID93 and COMMID97, which can be used to link households to the IFLS community-level data. Households that by 1997 had moved out of their 1993 village/municipality (MOVER97 = 2 or higher) have a missing value for COMMID97, since in 1997 they no longer lived in an area for which IFLS community data are available. There are two exceptions to this rule:

Twenty-seven households that moved from their 1993 location actually moved to another IFLS community and so can be linked to the CFS data for that community. For those households, COMMID97 is not missing.

Thirteen households relocated to the same area – an area where we decided to collect community data in 1997 since we found a cluster of IFLS households to be living there.

PTRACK

PTRACK contains a record for every person who appears in an IFLS1 or IFLS2 household roster. PTRACK contains 39,789 records, one for each of the 33,081 individuals listed in a 1993 household roster, and one for each of the additional 6,708 new members of origin and split-off households.

Within PTRACK, each observation is identified by PIDLINK. PTRACK contains a number of variables that will help establish the basic demographic composition of each IFLS wave and the availability of individual-level data from each wave. PTRACK indicates the tracking status of each 1993 household member and whether he/she was found in 1997. Variables indicate the our best guess of each person's age and sex, as well as information on marital status at each wave and the survey books for which data are available from each wave. PTRACK also indicates which books an individual answered in each wave. For example, it shows that in IFLS1 a certain woman respondent properly answered book 3 and did not answer book 4 because she was not married; by IFLS2 she was married and answered book 3 but failed to answer book 4, as she should have. Such information allows the analyst to calculate the number of observations in IFLS1 and IFLS2 and the number of panel observations for the various survey books.

²² Because administrative codes are revised quite frequently in Indonesia, we thought it important to provide the most recent codes we could obtain, in addition to the 1993 codes.

Finally, PTRACK shows the person's household and person IDs in IFLS1 and IFLS2. A person's household ID is often the same in both waves, but individuals who moved out of the 1993 household and were interviewed in a new household will have different household and person identifiers across waves. Individuals who were new household members in 1997 will not have HHID for 1993.

PTRACK does not provide information on individuals' locations. At the household level, that information is in HTRACK. For individuals who were new household members in 1997 (AR01a_97 = 5), the location information in HTRACK for 1993 is not necessarily the location where the new individual resided in 1993. To ascertain where a new household member lived in the past, data from module MG in book 3A should be used.

Tracking Changes in Characteristics across Survey Waves

Household Location. As noted above, in IFLS2 not all origin households were interviewed in their IFLS1 location. Variable MOVER97 identifies whether a household moved between 1993 and 1997 and, if so, whether it remained in any of the same administrative areas. MOVER97 also indicates, for split-off households, where the target respondent lived relative to his/her 1993 location.

Household Head. Sometimes the designated head in IFLS1 changed as of IFLS2. For example, if a couple divorced and the man moved out by 1997, the woman might be designated the household head in IFLS2. Households headed by elderly people in IFLS1 sometimes were headed by a son or son-in-law by IFLS2. When the household head changed, many of the relationships to the head changed as well.

Household Composition. Births, deaths, and individual moves affected household composition between the waves. The household roster variable AR01a_97 (which is also in PTRACK) indicates whether each individual was

Present in 1993 but had died before the 1997 interview (AR01a_97 = 0)

Present in both 1993 and 1997 (AR01a_97 = 1)

Present in 1993 but had moved out by 1997 (AR01a_97 = 3)

A 1993 household member interviewed in a split-off household in 1997 (AR01a_97 = 4)

A new member not present in 1993 (AR01a_97 = 5).

Marital Status. Some individuals who were single in 1993 had married by 1997; some who were married in 1993 were widowed or divorced by 1997.

Age or Year of Birth. In theory respondents interviewed in IFLS1 should have been three or four years older in 1997, depending on the time of year the interview took place in each wave. In Indonesia, as in many developing countries, however, not everyone knows his/her birthdate or age accurately. Therefore, reported birthdate across waves does not always match for a respondent, and there may be age discrepancies between waves (and even between books within a wave). The PTRACK file provides our best guess for age and birthdate.

Sex. For all but a few respondents, the reported sex matches across waves. The PTRACK file provides our best guess for sex in an attempt to resolve discrepancies.

Data Availability for Communities and Facilities: CTRACK and FTRACK

Files named CTRACK and FTRACK indicate what data are available for communities and facilities, respectively, in each survey wave.

CTTRACK contains a record for every IFLS community. Each observation is identified by COMMID93 and COMMID97, which match each other. In IFLS1, we visited 321 EAs, located in 312 different communities (as defined by the administrative boundaries of *desa/keurahan*) in which households were interviewed. Community-level information was collected for each of these 312 communities. In IFLS2 we again collected community-level information in each of these 312 communities, as well as in one additional community (COMMID97 = 5115), to which a number of IFLS1 households had moved by 1997. CTRACK indicates what community-level information is available for each survey year.

FTRACK contains information for each facility that was interviewed in either IFLS1 or IFLS2. Each observation is identified by FCODE. For each facility in FTRACK, variable STRATA defines the type of facility, INT93 indicates whether the facility was interviewed in 1993, and INT97 indicates whether the facility was interviewed in 1997. Some facilities were interviewed in both 1993 and 1997.

In 1997 we did not interview traditional practitioners. Therefore, none of the traditional practitioners interviewed in 1993 were reinterviewed in 1997. In both 1993 and 1997 we interviewed community health posts (*posyandu*). However, because the location and staffing of community health posts can change substantially over time, depending on the availability of the volunteers, we did not attempt to determine whether any health posts were interviewed in both years. None of the FCODES assigned to health posts in 1993 were assigned to health posts in 1997, and vice versa, although in fact some health posts may have been interviewed twice.

Although we collected community-level data in both years and reinterviewed facilities in 1997, there is no guarantee (nor was there any effort to ensure) that the respondents to the community and facility questionnaires were the same individuals.

Merging IFLS1 and IFLS2 Data for Communities and Facilities

The IFLS database can be used as a panel of communities and facilities. In both IFLS1 and IFLS2, data were collected at the community level from the leader of the community (book 1) and the head of the community women's group (book PKK). Data were also compiled from statistical records maintained in the community leader's office (book 2). The availability of these data makes it possible to examine change in community characteristics over time.

In IFLS2 and IFLS1-RR, variable COMMID identifies the IFLS communities, with an extension of 93 or 97 to indicate the source year. In the original IFLS1 data, communities were identified by variable EA.²³ To convert EA in the original IFLS1 data to COMMID93, the analyst may wish to use files in the COMMID93.ZIP file that accompanies the CFS data for IFLS2. COMMID93.ZIP contains two files: (1) a text file that has the crosswalk between EA and COMMID93, and (2) an .fmt file that contains formats for SAS users.

SAS users should use the following two statements to convert EA to COMMID93. Note that the full pathname for the location of the COMMID93.FMT file needs to be included in the %include statement.

SAS Code:

```
%include commid93.fmt;
```

²³ We changed community identifiers in IFLS2 because use of the three-digit EA code in both HHID97 and COMMID97 is misleading. In 1993, all IFLS households lived in one of the 321 IFLS EAs, so it was appropriate to identify both households and communities by EA. By 1997, some households had moved from their 1993 community. Their 1997 HHID still contains the three-digit EA code since it identifies the community from which they moved, but it does not identify the community of their current residence. Analysts should not merge households with community data on the basis of EA, for that would link movers to communities in which they no longer live.

```
length commi d93 $ 4;
commi d93=put(ea, commi d. );
```

Stata users should issue a series of “generate-replace” commands in combination with the COMMID93.TXT file that contains the EA-COMMID93 crosswalk:

Stata Code:

```
gen str4 commi d93 = XXXX if EA = =YYY
repl ace commi d93 = XXXX if EA = =YYY
```

In both IFLS1 and IFLS2, data were collected at the facility level from government health centers, private practitioners, community health posts, and schools (elementary, junior high, and senior high). In the original IFLS1, facilities were identified by the seven-digit character variable FACCODE. In IFLS2 and in IFLS1-RR, facilities are identified by the seven-digit character variable FCODE. Although the length is the same, the values of the variables differ. There is no point in trying to merge the original IFLS1 facility data with IFLS2 facility data simply by renaming one of the variables, because nothing should match, so any match is a mistake.

The file FACXWLK.ZIP contains several files that can assist the analyst in converting the FACCODE variable in IFLS1 data to an FCODE variable that will work with the IFLS2 data. When creating FCODE, remember that it needs to be a seven-digit character variable. In SAS, use a length statement (length FCODE \$ 7). In Stata, specify the creation of a string variable of length 7 (gen str7 FCODE). FACXWLK.ZIP contains the following files:

- FACXWALK.XPT: SAS transport file with FACCODE and FCODE
- FACXWALK.DTA: Stata version 6 file with FACCODE and FCODE
- FACXWALK.TXT: Text file with FACCODE and FCODE
- FACXWALK.FMT: SAS proc format that created format \$facxwalk.
- FACXWALK.SSD01: UNIX SAS files with FACCODE and FCODE.

As with COMMID, FACCODE can be converted to FCODE by using formats (SAS), by creating “if-then” statements (SAS), or by creating “generate-replace” statements (Stata).

In IFLS1, doctors and clinics were administered a different questionnaire from nurses, midwives, and paramedics. Because the questionnaires were different, the data were stored in different files. In IFLS2, all types of private practitioners received the same questionnaire and data are stored in the same files. To combine IFLS1 and IFLS2 data from private practitioners, the analyst should first combine the IFLS1 doctor/clinic data with the IFLS1 nurse/paramedic/midwife data.

Appendix A: Names of Data Files for the Household Survey

File Name	Contents	Level of Observation	No. Records
HTRACK	Household-level tracking across waves	Household	8116
PTRACK	Person-level tracking across waves	Individual	39789
BK_COV	Bk K Cover (Control Book)	Household	8116
BK_SC	Bk K Location and sampling	Household	7637
BK_KRK	Bk K Household characteristics	Household	7620
BK_AR0	Bk K Household size	Household	7620
BK_AR1	Bk K Household roster	Individual	39711
B1_COV	Bk 1 Cover (HH Economy)	Household	7608
B1_KS0	Bk 1 Consumption (1)-Misc	Household	7566
B1_KS1	Bk 1 Consumption (2)-Food	Food exp item	279942
B1_KS2	Bk 1 Consumption (3)-Non food mthly	Non food exp item	68094
B1_KS3	Bk 1 Consumption (4)-Non food ann	Non food exp item	52962
B1_KS4	Bk 1 Consumption (5)-Prices	Food item	37830
B1_PP1	Bk 1 Health facilities	Facility	83226
B2_COV	Bk 2 Cover (HH Bus, wealth)	Household	7616
B2_KR	Bk 2 Housing characteristics	Household	7600
B2_UT1	Bk 2 Farm business (1)-income	Household	7600
B2_UT2	Bk 2 Farm business (2)-assets	Asset	23310
B2_NT1	Bk 2 Non farm business (1)-income	Household	7600
B2_NT2	Bk 2 Non farm business (2)-assets	Asset	23625
B2_HR1	Bk 2 household Assets (1)-grid	Asset	83600
B2_HR2	Bk 2 household Assets (2)-debt	Household	7600
B2_HI	Bk 2 household non labor income	Income source	68400
B2_GE	Bk 2 household econ hardships	Shock	45600
B3A_COV	Bk 3A Cover (Individ Adult)	Individual	20529
B3A_DL1	Bk 3A Education (1)	Individual	19910
B3A_DL2	Bk 3A Education (2)	School	17385
B3A_DL3	Bk 3A Education (3)-grid	School	33284
B3A_DL4	Bk 3A Education (4)-expenses	Individual	8321

File Name	Contents	Level of Observation	No. Records
B3A_DLR1	Bk 3A Youth educ/emp (1)-summary	Individual	5333
B3A_DLR2	Bk 3A Youth educ/emp (2)-disruptions	School	26665
B3A_HR0	Bk 3A Individ assets (1)-screen	Individual	19910
B3A_HR1	Bk 3A Individ assets (2)-grid	Asset	75977
B3A_HR2	Bk 3A Individ assets (3)-debts	Individual	6907
B3A_HI	Bk 3A Individ non labor inc	Income source	179190
B3A_KW1	Bk 3A Marriage (1)-screen	Individual	19910
B3A_KW2	Bk 3A Marriage (2)-current	Individua	8503
B3A_KW3	Bk 3A Marriage (3)-history	Marriage	9832
B3A_PK1	Bk 3A HH decision making (1)	Individual	16994
B3A_PK2	Bk 3A HH decision making (2)	Decision	288898
B3A_PK3	Bk 3A HH decision making (3)	Status indicator	99304
B3A_BR	Bk 3A Pregnancy summary	Individual	2187
B3A_MG1	Bk 3A Migration (1)-birthplace	Individual	19910
B3A_MG2	Bk 3A Migration (2)-history	Migration event	16251
B3A_TK1	Bk 3A Work history (1)-screen	Individual	19910
B3A_TK2	Bk 3A Work history (2)-current job	Individual	13429
B3A_TK3	Bk 3A Work history (3)-history	Year	100453
B3A_TK4	Bk 3A Work history (4)-first job	Individual	14293
B3B_COV	Bk 3B Cover (Individ Adult)	Individual	20521
B3B_KM	Bk 3B Smoking	Individual	19892
B3B_KK	Bk 3B Self assessed health	Individual	19892
B3B_AK	Bk 3B Health insurance	Benefit	119352
B3B_MA1	Bk 3B Acute morbidity	Morbidity	437624
B3B_MA2	Bk 3B Morbidity-symptoms	Individual	19892
B3B_PS	Bk 3B Self-treatment	Treatment	79568
B3B_RJ1	Bk 3B Outpatient care (1)-use	Health facility	159136
B3B_RJ2	Bk 3B Outpatient care (2)-events	Treatment	4598
B3B_RN1	Bk 3B Hospitalization (1)-use	Health facility	99907
B3B_RN2	Bk 3B Hospitalization (2)-events	Treatment	476
B3B_PM1	Bk 3B Community participation (1)	Activity	159136
B3B_PM2	Bk 3B Community participation (2)	Individual	19892
B3B_PM3	Bk 3B Community participation (3)	Activity	238704
B3B_BA0	Bk 3B Non-HH mems (1)-parents	Individual	19892
B3B_BA1	Bk 3B Non-HH mems (2)-transfers	Individual	9144
B3B_BA2	Bk 3B Non-HH mems (3)-sibs (summary)	Individual	19892
B3B_BA3	Bk 3B Non-HH mems (4)-sibs (roster)	Sibling	53457
B3B_BA4	Bk 3B Non-HH mems (5)-sibs (transfers)	Individual	19892
B3B_BA5	Bk 3B Non-HH mems (6)-kids (summary)	Individual	19892

File Name	Contents	Level of Observation	No.Records
B3P_COV	Bk 3P(roxy) Cover (Individ Adult)	Individual	1655
B3P_KW1	Bk 3P(roxy) Marriage	Individual	1653
B3P_MG	Bk 3P(roxy) Migration	Individual	1653
B3P_DL1	Bk 3P(roxy) Education (1)	Individual	1653
B3P_DL3	Bk 3P(roxy) Education (3)-grid	School	4820
B3P_DL4	Bk 3P(roxy) Education (4)-expenses	Expense	1205
B3P_TK1	Bk 3P(roxy) Work (1)-screen	Individual	1653
B3P_TK2	Bk 3P(roxy) Work (2)-current job	Year	916
B3P_TK4	Bk 3P(roxy) Work (4)-first job	Individual	1217
B3P_PM1	Bk 3P(roxy) Commun partic (1)	Individual	1653
B3P_PM2	Bk 3P(roxy) Commun partic (2) activities	Activity	19836
B3P_KM	Bk 3P(roxy) Smoking	Individual	1653
B3P_KK	Bk 3P(roxy) Health status	Individual	1653
B3P_MA	Bk 3P(roxy) Acute morbidity	Morbidity	36366
B3P_RJ1	Bk 3P(roxy) Outpatient care (1)-use	Health facility	13224
B3P_RJ2	Bk 3P(roxy) Outpatient care (2)-events	Treatment	406
B3P_RN1	Bk 3P(roxy) Hospitalization (1)-use	Health facility	8315
B3P_RN2	Bk 3P(roxy) Hospitalization (2)-events	Treatment	56
B3P_BR	Bk 3P(roxy) Pregnancy summary	Individual	1653
B3P_CH0	Bk 3P(roxy) Pregnancy history (1)	Individual	606
B3P_CH1	Bk 3P(roxy) Pregnancy history (2)	Individual	83
B3P_CX	Bk 3P(roxy) Contraception	Individual	606
B3P_BA0	Bk 3P(roxy) Non HHM (1)-parents	Individual	1653
B3P_BA1	Bk 3P(roxy) Non HHM (2)-transfers	Individual	549
B3P_BA2	Bk 3P(roxy) Non HHM (3)-sibs (summary)	Individual	1653
B3P_BA3	Bk 3P(roxy) Non HHM (4)-sibs (roster)	Sibling	3349
B3P_BA4	Bk 3P(roxy) Non HHM (5)-sibs (transfrs)	Individual	1653
B3P_BA5	Bk 3P(roxy) Non HHM (6)-kids (summary)	Individual	1653
B3P_BA6	Bk 3P(roxy) Non HHM (7)-kids (roster)	Child	1413
B4_COV	Bk 4 Cover (Ever marr female)	Woman	6269
B4_KW1	Bk 4 Marriage	Woman	6160
B4_KW2	Bk 4 Marital history	Marriage	6785
B4_BR	Bk 4 Pregnancy summary	Woman	6160
B4_BA6	Bk 4 Non-HH members-children	Child	13322
B4_BX6	Bk 4 Non-HH members-children	Child	295
B4_BF	Bk 4 Breastfeeding (Panel resp.)	Woman	3984
B4_CH0	Bk 4 Pregnancy history (1)	Woman	6160
B4_CH1	Bk 4 Pregnancy history (2)	Pregnancy	5702

File Name	Contents	Level of Observation	No. Records
B4_CX1	Bk 4 Contraception (1)	Method	49280
B4_CX2	Bk 4 Contraception (2)	Woman	6160
B4_KL1	Bk 4 Contraceptive calendar (1)	Woman	6160
B4_KL2	Bk 4 Contraceptive calendar (2)	Month	661200
B5_COV	Bk 5 Cover (Child)	Individual	10415
B5_DLA1	Bk 5 Child's education (1)	Individual	10356
B5_DLA2	Bk 5 Child's education (2)-disruptions	Disruption	51778
B5_DLA3	Bk 5 Child's education (3)-history	School	7975
B5_MAA0	Bk 5 Child's health status	Individual	10356
B5_MAA1	Bk 5 Child's acute morbidity	Morbidity	269256
B5_PSA	Bk 5 Self-treatment	Treatment	41424
B5_RJA0	Bk 5 Out patient care-(1) use	Individual	10356
B5_RJA1	Bk 5 Out patient care-(2) services	Health facility	12880
B5_RJA2	Bk 5 Out patient care-(3) events	Treatment	2104
B5_RJA3	Bk 5 Out patient care-(4) vaccin	Individual	10356
B5_RNA1	Bk 5 Hospitalization - (1) use	Health facility	51780
B5_RNA2	Bk 5 Hospitalization - (2) events	Treatment	146
BUS_0	Bk US Health Assess (0)-HH summary	Household	7485
BUS_1	Bk US Health Assess (1)-Individ msr	Individual	37983
BEK	Bk EK Math/Bah Ind evaluations	Achievement test	22081

Appendix B: Names of Data Files for the Community-Facility Survey

File Name	Contents	Level of Observation	No. Records
BK1	BK1	BK1	313
BK1_A	BK1:A Destination	Destination	2817
BK1_B	BK1:B Electricity	Elec. Source	1818
BK1_D1	BK1:D1 Irrigation	Irrigation	1565
BK1_D2	BK1:D2 Extension Activity	Activity	284
BK1_D3	BK1:D3 Crop	Crop	939
BK1_D4	BK1:D4 Factory	Factory	1565
BK1_D5	BK1:D5 Cottage Industry	Cottage Industry	1565
BK1_E1	BK1:E1 Name Change	Name Change	13
BK1_E2	BK1:E2 Major Event	Major Event	3443
BK1_G	BK1:G Credit	Credit Inst.	2191
BK1_I	BK1:I History schools	School Level	939
BK1_J	BK1:J History Health Facility	Hlth Facility Type	1252
BK1_K	BK1:K Respondents	Respondent	576
BK1_PMKD	BK1:PMKD Activity	Activity	5008
BK1_RW	BK1:RW Neighborhood	Neighborhood	689
BK2	BK2: Community	BK2	312
BK2_HPJ	BK2:HPJ Price from retail	Item	9360
BK2_KA1	BK2:KA1 Environ. Conditions	Resource	1248
BK2_KA2	BK2:KA2 Land Ownership	Title	3432
PKK	PKK	PKK	310
PKK_H	PKK:H Local Prices	Item	12090
PKK_I	PKK:I History Schools	School	930
PKK_J	PKK:J History Health Facility	Facility	1240
PKK_KR	PKK:KR Resp Characteristics	Respondent	390
PKK_PM	PKK:PM Activity	Activity	3410
ADAT1	Adat1: Respond Characteristics	Adat1	304
ADAT2	Adat2: Traditions	Time Period	608
ADAT_AP1	Adat: AP1-Marriage gifts	Respondent	1216

File Name	Contents	Level of Observation	No. Records
PM	Community participation	Community	303
SAR	Service Availability Roster	SAR	15260
PUSK	PUSK	Puskesmas	922
PUSK_B1	PUSK:B1 Activity/Service	Activity/Service	921
PUSK_C1	PUSK:C1 Service	Service	35919
PUSK_C2	PUSK:C2 Referral Facility	Facility	4605
PUSK_C3	PUSK:C3 Laboratory Test	Test	10131
PUSK_D	PUSK:D Employee	Employee	6559
PUSK_E1	PUSK:E1 Equipment	Equipment	20262
PUSK_E2	PUSK:E2 Supplies	Supply	11973
PUSK_F	PUSK:F Medicines	Medicine	25788
POS	Posyandu	Posyandu	619
POS_B1	Posyandu: B1-Hlth services	Hlth service	6190
POS_B2	Posyandu: B2-FP services	FP service	7428
POS_C	Posyandu: C-Personnel	Worker	4333
POS_D	Posyandu: D-Hlth equipment	Equipment	8047
POS_H	Posyandu: H-Local prices	Item	24141
PRA	PRA	Priv Practice	1832
PRA_B1	PRA:B1 Opening and Closing Time	Day	12824
PRA_B2	PRA:B2 Service Availability	Service	73280
PRA_B3	PRA:B3 Referral Facility	Facility	7295
PRA_B4	PRA:B4 Laboratory Tests	Tes	14656
PRA_C1	PRA:C1 Health Equipment	Equipment	36640
PRA_C2	PRA:C2 Health Supplies	Supply	36640
PRA_D1	PRA:D1 Stock of Meds	Medicine	49464
SD	SD: School	School	964
SD_B2	SD:B2 Schools sharing building	School Type	5784
SD_B3	SD:B3 Schools sharing complex	School Type	2250
SD_B4	SD:B4 Facility type	Facility type	8676
SD_C	SD:C Teacher	Teacher	1927
SMP	SMP: School	School	945
SMP_B2	SMP:B2 Schools sharing building	School Type	5670
SMP_B3	SMP:B3 Schools sharing complex	School Type	1152
SMP_B4	SMP:B4 Facility type	Facility type	8505
SMP_C	SMP:C Teacher	Teacher	1890

File Name	Contents	Level of Observation	No. Records
SMU	SMU: School	School	618
SMU_B2	SMU:B2 Schools sharing building	School Type	3708
SMU_B3	SMU:B3 Schools sharing complex	School Type	900
SMU_B4	SMU:B4 Facility type	Facility type	5562
SMU_C	SMU:C Teacher	Teacher	1235

Appendix C: Module-Specific Analytic Notes

This appendix presents detailed notes about IFLS2 data from the household survey that may be of interest to analysts who will use the data.

Book K: Control Book and Household Roster

Book K recorded whether a household was found and interviewed and the location of households that were found. If the household was interviewed, information was collected on the composition of the household and on basic housing characteristics that the interviewer could observe.

Cover (BK_COV)

Some respondents listed on the cover page were not household members. If the household was not found, a neighbor or other community member most likely provided information about the household's whereabouts. In some cases the household was found and interviewed, but the residents were infirm or otherwise unable to answer for themselves, so someone who knew them well answered. In some cases the respondent listed on the cover lived in the household in 1993 but not in 1997. In these cases the respondent's PID number is given, since the roster will provide information on that person. In a few cases a person younger than age 15 provided information for book K.

Module AR (BK_AR0, BK_AR1, BK_AR3)

1. For origin households much information from the 1993 household roster was preprinted on the 1997 roster so that interviewers would know whom they were looking for and to obtain updated information on all 1993 household members. The preprinted variables include PID97 (AR00/PERSON in IFLS1), AR01, AR02, AR00id (PIDLINK), AR07, AR08, AR08a, AR01b, AR01c, and AR01d. Information now in the data set is not necessarily the information that was preprinted. For certain variables, such as birthdate, interviewers often updated the preprinted information in the field. Therefore, the birthdate reported in AR08 may not match the birthdate in IFLS1 data.
2. Variable AR01a indicates the household member's status in the 1997 household:
Origin households:
 - 0 = 1993 member deceased in 1997
 - 1 = 1993 member still in 1997 household
 - 3 = 1993 member who had left by 1997
 - 5 = 1997 member not present in 1993 (new member)
 Split-off households:
 - 4 = member of origin household interviewed in a new household in 1997
 - 5 = member of household in 1997 but not in any origin household in 1993
3. In the fielded version of the survey, variables AR01c and AR01d indicated whether a respondent should be treated as a panel or new respondent in books 3 and 4. We have replaced the preprinted information with the actual treatment of the respondent in the field (PANEL3 and PANEL4). We have also included variables that indicate whether eligible respondents completed books 3A, 3B, 4, Proxy, and 5.
4. Variables AR10, AR11, AR12, and AR14 provide the roster line number (PID97) of an individual's father, mother, caretaker (for children), and spouse (for married respondents), if

they were members of the household. Because the preprinted rosters contained all 1993 household members, an individual's father, mother, caretaker, or spouse sometimes had a PID in the roster but was not a current member of the household. Such cases were not handled consistently in the field. Sometimes line numbers were filled in; other times code 51 (not in household) was entered. To prevent confusion, we have left line numbers in the data only if both the respondent and the respondent's relative were current members of the household.

5. IFLS2 added new questions on whether respondents had worked in the past 12 months and their salary if they had worked. These questions provide useful information on the activities of former household members and of individuals whom we failed to interview in 1997. However, earnings appear to have been significantly underreported. For example, if household labor income is calculated by summing AR15b for current household members, a number of households appear to have no earnings.
6. In some households, the person who answered book K noted errors in the preprinted information on 1993 household composition. In 383 cases the respondent said that a person listed on the preprinted roster had not been living in the household in 1993. In 10 cases the respondent said that a person not on the roster was a household member in 1993. It is not clear whether the preprinted information or the 1997 respondent was wrong.

Module KKK (BK_KRK)

Module KKK was for interviewers to fill out at the end of the first interview, based on observations of the household. We believe that some interviewers had the respondent provide information such as the number of rooms in the house and the size of the house in square meters.

Book 1: Expenditures and Knowledge of Health Facilities

Book 1 was typically answered by a female respondent, either the spouse of the household head or another person most knowledgeable about household affairs. One module recorded information about household expenditures²⁴ and about quantities and purchase prices of several staples. The other module probed the respondent's knowledge of various types of public and private outpatient health care providers.

Cover (B1_COV)

A few respondents were younger than age 12 because it was determined that no available older person would be a better respondent.

Module KS (B1_KS0, B1_KS1, B1_KS2, B1_KS3, B1_KS4)

1. Some households reported little or no food expenditures. We believe that those data are correct because notes indicated that the household was a special case. For example, the food expenditures of a household that operates a *warung* are impossible to separate from food expenditures for the warung. Another household had only member, a student who took all his meals at the university, where food was included in the cost of tuition.

²⁴ IFLS1 and IFLS2 included similar topics and reference periods for expenditures. For a subset of items, IFLS1 asked whether the reported expenditures pertained only to the individual answering the questions or to the entire household. That question was dropped in IFLS2 because the whole module was supposed to apply to the entire household. The expenditure module was a shortened version (about 40 minutes) of the three-hour budget expenditure survey conducted by BPS.

2. Expenditure questions dealt with different reference periods: weekly, monthly, and yearly. 62
Calculation of total expenditures requires standardizing on one reference period.
3. Questions KS13a–KS15 attempted to obtain food prices for standard units of measure.
Respondents had two chances (KS14 and KS14b). Some respondents would not provide the price for a standard unit.

Module PP (B1_PP1, B1_PP2)

In answering the module's questions about sources of health and family planning facilities, the respondent could mention any facility in any location, near or far. PPTYPE covers 11 types of facility, chosen to cover the types of services typically available. The facility types listed do not necessarily match respondents' definitions of facilities. For example, respondents did not always know whether a hospital was public or private, or whether a provider was a doctor versus a paramedic or a nurse versus a midwife.

Book 2: Household Economy

Book 2 was typically answered by the household head or the head's spouse. Some modules asked about household businesses (farm and nonfarm), nonbusiness assets, and nonlabor income. Other modules collected information about housing characteristics and economic shocks experienced by the household in the previous five years.

Module KR

1. Respondents had difficulty answering question KR5, which asked homeowners to estimate the rental price they could get if they were to rent their home.
2. Question KR26 asked whether the household had a *Kartu Sehat* (health card). The term was intended in its precise meaning — a card given by the community leader ostensibly to needy households that entitles members to free or subsidized health care at the public health center. Some respondents may have interpreted the term generically and may have reported possessing a *Kartu Sehat* when they meant simply an insurance identification card.

Module UT

1. UT04 and UT05 asked about other owners of the farm business. The list from which respondents could choose was not exhaustive. Responses recorded in the Other category may not capture all cases where a business was owned with children-in-law or with ex-spouses (or family of an ex-spouse). Since those relationship categories were not listed, the respondent had to report them specifically in the Other category.
2. UT05 and UT06 respectively report who in the household owned the business, and what fractions were owned by husband and wife. UT05 sometimes identified both respondent and spouse as owners, but UT06 recorded only one of them as owner. In other cases, the spouse was not identified as an owner in UT05, but a fraction of ownership was reported in UT06. Reports of fractions owned by husband and wife do not always add up as expected. Sometimes husband and wife are not the only owners in the household, but their shares add up to 100%. Other times the husband and wife are the only owners, but their shares add up to less than 100%.

Module NT

1. NT04 and NT05 asked about other owners of nonfarm businesses. The list from which respondents could choose was not exhaustive. Responses recorded in the Other category may not capture all cases where a business was owned with children-in-law or with ex-

spouses (or family of an ex-spouse). Since those relationship categories were not listed, the respondent had to report them specifically in the Other category. ⁶³

2. NT05 and NT06 respectively report who in the household owned the business, and what fractions were owned by husband and wife. The answer to NT05 sometimes identified both respondent and spouse as owners, but NT06 recorded only one of them as owner. In other cases, the spouse was not identified as an owner in NT05, but a fraction of ownership was reported in NT06. Reports of fractions owned by husband and wife do not always add up as expected. Sometimes husband and wife are not the only owners in the household, but their shares add up to 100%. Other times the husband and wife are the only owners, but their shares add up to less than 100%.

Module HR

HR10 asked who owned household or “nonbusiness” assets, and HR12 asked what fractions were owned by husband and wife. HR10 sometimes identified both respondent and spouse as owners, but HR12 recorded only one of them as owner. In other cases, the spouse was not identified as an owner in HR10, but a fraction of ownership was reported in HR12. Reports of fractions owned by husband and wife do not always add up as expected. Sometimes husband and wife are not the only owners in the household, but their shares add up to 100%. Other times the husband and wife are the only owners, but their shares add up to less than 100%.

Module HI

Module HI asked about nonlabor income, and space was provided to record types of nonlabor income not listed. When the answers recorded in that Other category were translated, we learned that some respondents had reported income from working and from transfers, which should have been reported in the module TK and module BA, respectively, and in fact may have been reported there as well.

Module GE

Some of the dates respondents reported for calamitous events (GE02) may not be precise. A sickness, crop loss, or business failure might have occurred over a period of months.

Book 3A: Adult Information (part 1)

Book 3A asked all household members 15 years and older about their educational, marital, work, and migration histories. In addition, the book included questions on asset ownership and nonlabor income, household decision-making, fertility preferences, and (for women 50 and older) cumulative pregnancies.

Module DL

1. Several DL questions pertained to schooling, including the date of leaving school and dates various EBTANAS tests were taken. We would expect the usual schooling sequence (e.g., start of school around age 6, elementary-level EBTANAS test six years later) to be reflected in the DL responses. However, a logical sequence does not appear for some respondents. In particular, respondents seemed to have difficulty reporting dates of entering school. Dates of EBTANAS tests, often taken directly from an EBTANAS score card, are believed to be more reliable.
2. When asked about the school level currently attended, respondents who attended *Madrasah* often reported that school type rather than the actual level (elementary or junior high school). For those respondents we substituted the appropriate level and entered “private Islam” in response to the question on the administrative type of the respondent’s school.

3. The EBTANAS scores in variable DL16d are not necessarily comparable across the country. 64
Local administrators had some control over the contents of the EBTANAS tests in their area until standardized versions were adopted. Standardized EBTANAS tests were implemented at the elementary level in the early 1990s and at the junior and senior high school levels in the mid-1990s. We recommend that analysts include controls for region when pooling EBTANAS scores across regions.
4. Whenever possible, interviewers recorded EBTANAS scores from the EBTANAS score card. Otherwise, the interviewer had to rely on the respondent's report. Generally EBTANAS scores have two digits to the right of the decimal and one digit to the left. Respondents had difficulty accurately recalling the two digits to the right of the decimal point. Heaping of responses on the *special codes* of 96–99 occurred. Some of those numbers may be valid responses; it is difficult to tell. Rather than creating two X variables (one for the number to the left of the decimal, one for the number to the right), we created only one X variable, indicating whether the respondent was able to provide any portion of the score. If the second two numbers are 96–99, we created a flag that warns analysts to inspect the scores and decide whether a number such as 98 to the right of the decimal point represents a valid score or an imprecise answer.
5. The questionnaire listed all the EBTANAS subject areas that we were able to identify. In fewer than 20 cases, interviewer's notes in the CP module at the end of the book indicated that a respondent had taken a test in an unlisted subject, such as accounting or a religious subject. This occurred more often at the senior high school level, where the curriculum varies more than in other levels. When the CP data note such an exception, we list the HHID, PID, subject, and score in the Special Cases list. Analysts may incorporate that information as they choose.
6. A respondent's total EBTANAS score did not always equal the sum of the scores for the component tests. Perhaps not all the subjects on which the person was tested were listed on the form, or perhaps the respondent forgot some component scores but remembered the total score.
7. Data from *interviewer checks*, where previous responses were recorded to ensure proper skip patterns (e.g., respondent's age, timing of schooling, and whether the respondent is panel or new), showed some errors, about 25–75 cases per skip. We generally did not correct skip patterns because of their complicated nature and the risk of overwriting data, albeit data that may have been collected in error.

Module DLR

Although respondents were questioned about absences from school lasting at least four weeks, some respondents reported absences of shorter durations. We did not remove those data.

Module HR

The notes about module HR in book 2 apply to book 3A as well.

Module KW

1. In reporting the value of the dowry at the time of the wedding (KW12b and KW13), some respondents cited old units of currency. Rather than trying to convert the values to the Indonesian rupiah (without knowing the proper conversion rate), we have provided codes that indicate the currency that was specified.
2. Questions KW14a–g asked both husband and wife about decisions on where and with whom to live after marrying. Look Ups checks revealed that the responses were not always consistent. We generally made no corrections because it wasn't clear which answer was correct. To investigate these inconsistencies further, the analyst could compare the information in module MG.

3. The “current” spouse was not always the same as the “latest” spouse if the respondent had had⁶⁵ two wives at one time and was still married to the wife he married *before* marrying the wife from whom he is now divorced.

Module BR

A woman’s total number of pregnancies reported here is not always consistent with the number of her offspring reported elsewhere. For example, some women reported fewer non-coresident sons in module BR than they reported in module BA. Perhaps the BA report includes someone who was not a biological child. Or, a son may have been inadvertently omitted from the BR report.

Module PK (B3a_PK1, B3a_PK2, B3a_PK3)

Some respondents to this module’s questions about household decision-making practices erroneously indicated that a particular topic was not applicable to them, whereas it was clear that a decision had been made. For example, a couple declared that the question of who decided about contraception was inapplicable, but they reported not using a contraceptive method. Similarly, another couple thought the question about deciding whether the woman should work was inapplicable, but the woman does not work.

Module MG

1. In designing IFLS2 we decided to ask both panel and new respondents their full retrospective migration history, rather than to ask it only of new respondents. Unfortunately the skip pattern in the questionnaire directed panel respondents to begin the retrospective history at question MG19b (moves since the age of 12), rather than at the question on place of residence at age 12. Thus, although we know the history of moves since age 12 for all respondents, for panel respondents we do not know the location of residence at age 12. That information should be available in IFLS1.
2. For respondents who reported moves in module MG, the last place to which they report moving should match the current residence recorded in module SC for the household. In some cases the two locations do not match.

Book 3B: Adult Information (part 2)

Book 3B emphasized current rather than retrospective information. Separate modules addressed insurance coverage, health conditions, use of inpatient and outpatient care, and participation in community development activities. Another module asked in detail about the existence and characteristics of non-coresident family members (parents, siblings, and children) and about whether money, goods, or services were transferred between these family members during the year before the interview.

Module KM (B3B_KM)

Question KM03 asked respondents whether they smoked filtered or unfiltered cigarettes. A number of respondents who reported smoking self-rolled cigarettes did not report whether the cigarettes were filtered or not. Since self-rolled cigarettes are presumably rolled without filters, we created a new category to so indicate.

For some respondents, the age at which they reported starting to smoke (variable KM10) was much greater than their current age. Where KM10 was 61 or higher and the respondent was younger than 61,

we assumed that the respondent had reported the year smoking began rather than his or her age. In those cases, we changed the data to reflect the respondent's age during the reported year. In 21 other cases this assumption did not appear warranted, and we left the inconsistency. 66

Module PS (B3B_PS)

In reporting self-treatment with various kinds of medicines, about 55 respondents reported medicines in the Other category that they had received from providers. Those medicines may also have been reported in module RJ or RN. To permit checks for double-counting, we indicate a PSTYPE code of G for the applicable Other medicines. Analysts can use the codes to compare medicines in RJ or RN and judge whether the same medicines were reported twice.

Module BA (Parent) (B3B_BA0, B3B_BA1)

Data are provided about the survival status and characteristics of parents living outside the household, and about transfers of money, goods, or services between the respondents and those parents.

1. BA data about parents' survival status and residence do not always agree with information in module AR. It is difficult to ascertain which module is correct. One legitimate reason for discrepancies is that AR10 and AR11 explicitly asked about the respondent's *biological* parents, whereas BA questions did not specify. Therefore, parents reported as dead in AR10 or AR11 could be biological parents, and the apparently conflicting data on parental characteristics and transfers in module BA could refer to step- or adoptive parents.
2. Some PIDs for persons identified in BA04a as parents of the respondent conflict with other information suggesting the impossibility of that particular relationship. Analysts should not assume that the line numbers in BA04a are completely accurate.
3. When asked about a parent's age, over 300 respondents reported a figure over 100. We have not changed these data, although it seems unlikely that so many respondents would have parents of that advanced age. Analysts may wish to cross parent's reported age against respondent's age to identify cases where the parent is implausibly older than the respondent.
4. Questions BA10m and BA10p established the applicability of questions about transfers. Transfer questions were not supposed to be asked about parents who had been dead for more than one year or about parents living in the household. However, the logic and the formatting of these questions were complicated. In a number of cases, respondents whose parents lived in the household reported transfer information about those parents. We have corrected BA10m and BA10p to indicate the parents' "correct" status, but we did not change BA10A or delete the erroneously collected transfer data.

Module BA (Sibling) (B3B_BA3, B3B_BA4, B3B_BA5)

Data are provided about the characteristics of non-coresident siblings and about transfers of money, goods, or services between respondents and those siblings.

1. For respondents who reported siblings in 1993, we preprinted the name, age, and sex of all siblings alive in 1993. In 1997, interviewers were supposed to use these preprinted sibling rosters to collect data on the same siblings (as well as others who had been missed, such as those younger than 15 in 1993 but 15 or older by 1997). Some preprinted sibling rosters were not used in the field, even though the respondent was interviewed. In those cases, we created variable PPSIB to indicate whether a preprinted sibling roster
 - 1 = existed and was used
 - 2 = existed but was not used
 - 3 = did not exist.

The same issue affects file B3P_BA3. Approximately 3% of cases are coded as 2 (preprinted⁶⁷ roster not used). There is imperfect agreement between PPSIB and the skip pattern questions that indicate whether a preprinted roster was used. We have not tried to resolve these inconsistencies and have more confidence in PPSIB. PPSIB is critical for linking 1993 data on siblings with 1997 data on siblings (see also BA30A_93).

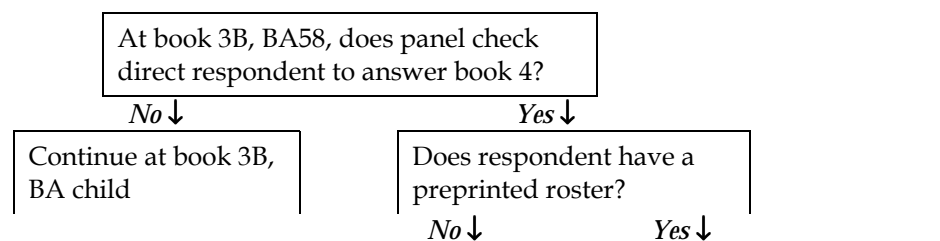
2. Where a preprinted sibling roster was used, variable BA30A_93 identifies the line number of the sibling in the 1993 data. Where a preprinted sibling roster was not used, BA30A_93 is missing for the respondent. There are some respondents for whom 1993 sibling data exist but a preprinted sibling roster was not used in 1997. In order to match sibling data from 1993 to data on the same sibling in 1997 for those respondents, analysts will have to use characteristics such as age and sex, since there is no guarantee that siblings were listed in the same order in both years. This issue also affects file B3P_BA3.
3. For a small number of respondents with preprinted sibling information, the 1997 interview indicated that the same sibling had been listed twice. To identify those cases, we have created the variable SAME_SIB. If a sibling is listed twice, for each listing SAME_SIB indicates the line number of the other record for that sibling. For example, if sibling 1 and sibling 2 are really the same person, SAME_SIB = 2 for sibling 1 and SAME_SIB = 1 for sibling 2. If SAME_SIB is missing, there is no evidence of a duplicate listing. This issue also affects file B3P_BA3.
4. In 14 cases, the interviewer or editor noted that the person as sibling in the preprinted list was not the respondent's sibling. The variable NOT_SIB flags those cases.

Module BA (Child) (B3B_BA6; see also B3P_BA6, B4_BA6, B4_BX, B4_CH1)

Data are provided about the characteristics of non-coresident children and about transfers of money, goods, or services between respondents and those children.

In IFLS1 all respondents answered questions about non-coresident children in book 3. As a result, women age 15–49 had to answer questions about their children in book 3 and again in book 4. In IFLS2 this protocol was changed to shorten the interview for women of reproductive age. Briefly, duplicate BA (child) modules were provided in IFLS2 books 3 and 4. Women 50 and older only had to answer questions in book 3, BA (child), and women age 15–49 only had to answer the questions in book 4, BA (child).

Skip Patterns. In IFLS2, book 3B, module BA (child), was administered to new respondents age 50 and older and to women who were panel respondents to book 3 who were 54 or older (i.e., too old to have received book 4 in IFLS1). Book 4, module BA (child), was administered to new respondents age 15–49 and panel respondents who had answered book 4 in IFLS1. For panel respondents to book 4 who had a preprinted child roster, questions about children who were alive as of 1993 were asked on the preprinted BA (child) roster (inserted in book 4), and questions about children born after 1991 were asked in module CH. For panel respondents to book 4 who did not have a preprinted child roster and for new respondents to book 4, questions about children were asked in module CH, which starts by listing all pregnancies and therefore all children ever born. Beginning with question CH28a, the questions in module CH are the same those in module BA. The figure below diagrams the skip patterns followed for women respondents.



Go to CH for pregnancy history (after CH27, questions duplicate BA child). Continue to BX for adopted or stepchildren.

Insert roster in book 4, BA (child), and update info on all children alive in 1993. Continue to CH for any children born after 1991.

68

Linking Children in 1997 Rosters to Their IFLS1 Data. To facilitate linking data on children in the 1997 rosters to data on those same children in 1993, we have provided the following variables:

AR00_93 (1993 household roster number)

BA63A_93 (line number in 1993 BA roster)

CH05_93 (column number in 1993 pregnancy roster).

Children listed in the 1993 *household* roster (for whom AR00_93 is not missing) will not be listed in the 1993 *non-coresident child* roster (therefore, BA63A_93 will be missing). Likewise, children listed in the 1993 non-coresident child roster (for whom BA63A_93 is not missing), will not be listed in the 1993 household roster (therefore, AR00_93 will be missing).

Lost/Missing Preprinted Child Rosters. For respondents who reported children in 1993, we preprinted the name, age, and sex of all children alive in 1993. In 1997 interviewers were supposed to use these preprinted child rosters to collect data on the same children. In some cases a preprinted child roster was created but was not used in the field, even though the respondent was interviewed. In these cases, we created the variable PPCHILD to indicate whether a preprinted child roster existed and was used, existed but was not used, or did not exist. This same issue affects file B3P_BA6.

Duplicate Listings for Children. For a small number of respondents with preprinted child information, the 1997 interview indicated that the same child had been listed twice. To identify these cases, we have created the variable SAMEKID. If a child is listed twice, for each listing SAMEKID indicates the line number of the other record for that child. For example, if child 1 and child 2 are the same person, SAMEKID = 2 for child 1 and SAMEKID = 1 for child 2. If SAMEKID is missing, there is no evidence of a duplicate listing. This issue also affects file B3P_BA6.

Ages of Non-Coresident Children. The instructions at book 3B, BA58, specified that new respondents and panel respondents without a preprinted child roster should list only non-coresident children age 15 and older. To be consistent with IFLS1, the instructions should have required the listing of all non-coresident children, regardless of age.

Book 3B child rosters for new respondents and panel respondents without a preprinted roster do not include non-coresident children younger than 15. For women, information from BR can be combined with information from the non-coresident child roster to ascertain the number of non-coresident children younger than 15 (BR addresses all non-coresident children; BA, non-coresident children 15 or older). Some men respondents may have non-coresident children younger than 15 who were born to a woman other than the respondent's current wife. There is no way to ascertain the number of these children.

Survival Status of Children: For a small number of cases, BA data indicate that a child who was in the household roster in 1993 had died by 1997, but other modules suggest that we successfully tracked that child to a new household in 1997. We have not changed respondents' reports on the survival status of their children where additional evidence suggests that the reported status is incorrect.

Pregnancy History and Children. The instructions at book 4, CH27x, skip panel respondents out of the CH module. The skip was correct for panel respondents who had a preprinted child roster because they had already answered questions about children who were alive as of 1993 on the preprinted BA roster. The skip was incorrect for panel respondents who did not have a preprinted child roster. Lacking the roster, we should have asked about the characteristics of children and transfers to and from them. Fortunately, few cases are affected. For most panel women without a preprinted roster, the roster did not exist because the woman did not have any children as of IFLS1. Such women were unlikely to

have any non-coresident children as of 1997. For only about 48 women does it appear that a preprinted⁶⁹ child roster existed but was not used.

Book 4: Ever-Married Woman Information

Administered to all ever-married women age 15–49, and to panel respondents who had answered book IV in 1993, book 4 collected retrospective life histories on marriage, children ever born, pregnancy outcomes and health-related behavior during pregnancy and childbirth, infant feeding practice, and contraceptive use. The marriage and pregnancy summary modules replicated those included in book 3 so that women who answered book 4 skipped these modules in book 3. Similarly, women who answered questions about non-coresident family in book 4 skipped that module in book 3. A separate module asked married women about their use of contraceptive methods on a monthly basis over the previous 5 to 10 years.

Module KW

The notes about module KW in book 3A apply to book 4 as well.

Module BA (Child) (B4_BA6)

For panel respondents who were to receive book 4, module BA (child), rather than its book 3 counterpart, we preprinted the name of the woman's youngest child as of 1993 at the bottom of the preprinted child roster. Two purposes were served. (1) If the youngest child was age 8 or younger in 1997 (and therefore 4 or younger in 1993), we were alerted to update IFLS1 information on breastfeeding, to obtain the duration of breastfeeding for children who might have still been breastfeeding in 1993. (2) The name of the youngest child provided an anchor for asking women to update their IFLS1 pregnancy information—about any pregnancies *following* the pregnancy that produced the youngest child reported by the respondent in 1993.

Module BF (B4_BF):

For children being breastfed at the time of IFLS1, this module provides updated information. The preprinted child roster for these children's mothers listed the name of the youngest child the mother reported in 1993. For children younger than 8 (that is, younger than 4 in 1993), data include the duration of breastfeeding, in case those children were still being breastfed at the time of the 1993 interview. In this module, "youngest child" means youngest child as of 1993. A few children are reported as being younger than 4 in 1997. It is unclear whether they were born after the 1993 interview or the age is wrong.

Module CH (B4_CH0, B4_CH1)

Variables CH01ab, CH01ac, and CH02a summarize pregnancies since the last interview for panel respondents. CH02a should equal the sum of CH01ab and CH01ac. Variables CH01Ba, CH01bb, CH01bc and CH02b summarize all pregnancies for new respondents. CH02b should equal the sum of CH01Ba, CH01bb, and CH01bc. CH03 indicates the number of pregnancies about which information should be collected, and it should equal either CH02a or CH02b. In about 20 cases, one or more of these arithmetic relationships does not hold. It is difficult to know which variable is in error.

Book 5: Child Information

Book 5 collected information about children younger than 15. For children younger than 11, the child's mother, female guardian, or caretaker answered the questions. Children between the ages of 11 and 14 were allowed to respond for themselves if they felt comfortable doing so. The five modules focused

on the child's educational history, morbidities, self-treatment, and inpatient and outpatient visits. Each 70 paralleled a module in the adult questionnaire (books 3A and 3B), with some age-appropriate modifications. For example, the list of acute health conditions specified conditions relevant to younger children.

Cover (B5_COV)

Sometimes book 5 was answered by an older sibling. Occasionally the older sibling was younger than age 15. Sometimes book 5 was answered by someone who was no longer in the household – for example, an aunt who had lived in the household in 1993, was no longer living in the household in 1997, but was deemed the most knowledgeable source of information for the child. In those cases the aunt's PID number from the roster is in the book 5 cover data (even though she is no longer a household member) since the roster contains information about the aunt's characteristics.

Module DLA (B5_DLA1)

1. Regarding the age at which the respondent entered elementary school, in about 100 cases the age reported (or calculated using information in DL03 and elsewhere) is less than 4. In Indonesia, most children enter elementary school at age 6 or 7. Though the less-than-4 data seem incorrect, we have left them, having no basis for making corrections. Some respondents may have interpreted the question as referring to the age of entering preschool.
2. DLA11 and DLA12 ask about hours worked per week on school days and per day on nonschool days. For some respondents relatively large numbers of hours were reported per week (although for fewer than 25 respondents was it more than 40). Some interviewers or respondents may have reported the total hours worked per week on nonschool days instead of per day, as asked.
3. For questions DLA23a–e, interviewers recorded EBTANAS scores from the EBTANAS score card whenever possible. Otherwise, the interviewer had to rely on the respondent's report. Generally EBTANAS scores have two digits to the right of the decimal and one digit to the left. Respondents had difficulty accurately recalling the two digits to the right of the decimal point. Heaping of responses on the *special codes* of 96–99 occurred. Some of those numbers may be valid responses; it is difficult to tell. Rather than creating two X variables (one for the number to the left of the decimal, one for the number to the right), we created only one X variable, indicating whether the respondent was able to provide any portion of the score. If the second two numbers are 96–99, we created a flag that warns analysts to inspect the scores and decide whether a number such as 98 to the right of the decimal point represents a valid score or an imprecise answer. The flag variables are named PROB23a-PROB23e. Nearly 150 cases have at least one DLA23* variable flagged as a problem.
4. In questions DLA29, DLA32, DLA33, respondents were asked about absences from school lasting at least four weeks. Fewer than 25 respondents reported absences of shorter durations. We did not remove those data.

Appendix D: Special Cases

This appendix lists records with unique characteristics that could not be reflected in the electronic data. Analysts may want to handle these cases differently from others of their type. The “CP notes” cited here refer to notes made in the CP module, located at the end of nearly every questionnaire book, which asked the interviewer to record the conditions of the interview, the respondent’s level of attention, and any other information that might clarify or explain the respondent’s answers.

Book 2, Module NT. Two respondents to questions NT07 and NT09 gave values of rupiahs per day rather than rupiahs for the last 12 months. According to the CP notes:

HHID = 330716826 (NT09 = 3000 Rp per day)

HHID = 330716806 (NT07 = 2000 Rp per day)

Book K, Module AR. A CP note says that for HHID = 331717402 and PID = 2 the value of AR15B should be 14,000 rupiahs/10 days. We did not change AR15B because data from the TK module indicated that the respondent only worked 2 weeks during the previous year.

Book 5, Module DLA. PIDLINK = 306240009 has unusual data in B5_DLA datasets. In DLA07/08, the child reports now being in school. DLA03 says the respondent entered school in 7/1997. DLA30 = 4 (# of absences in past 5 years). In the DLA2 dataset the same respondent reported not having been in school for the entire years 1993, 1994, 1995, and 1996 with the reason “Could not afford.” So it appears that the child should have entered school in 1993 (is 11 years old at the interview date) but could not because of the cost. The child interpreted this as absences rather than as not entering school until age 11.

Book 3A, Module DL. Question B3A_DL3 did not list all possible component tests for the EBTANAS exam. Occasionally the interviewer entered a CP note about additional EBTANAS exams a respondent had taken. The following table reproduces the relevant CP notes. The user should assume that the total EBTANAS score (in DL16E) reflects all tests taken by the respondent.

HHID	PID	Row	Score	CP Note
317308211	4	3	4.4	Grade for cost accountancy lesson and 2.95 is grade for finance accountancy.
317308614	3	3	6.2	Grade for secretary lesson
317510018	4	3	4.45	Grade for finance accountancy, and 4.50 is grade for cost accountancy.
320411005	6	3	4.8	Grade for “textile finishing technology” lesson
327114812	6	3	2.8	Grade for “secretary productivity” lesson
990316906	1	2		Lesson Quran Hadist of Islamic culture 6.07 (total “Danem” is 36.36)
340220011	3	3		IPA :5.40 , Jasa Boga ; 5.20
HHID	PID	Row	Score	CP Note

340220307	4	3		Value of Akutansi Biaya is 6.05 and Aku Tansi Keuangan is 5.25 didn't added on Sr High
351523719	4			Physics 5.30, Biology 6.00, Alquran 6.90, Fiqih 7.60, Bahasa Arab 6.90, so total of nem was 65.30
351523720	2			Physics 5.50, Biology 7.00, Alquran 6.60, Fiqih 6.10, Bahasa Arab 4.60, so total of nem was 61.00
352524812	7	N	3.4	Grade for general mechanics lesson
510526930	4			The point of sociology lesson and anthropology were joint together, 5.50
520328207	8			Add subjects: Qur'an Hadist 7.80, Fiqh 4.85, Arabic language, 6.00, Physics 3.20
520328210	4			Add subjects: Qur'an Hadist 4.70, Fiqh 3.30, Arabic language 3.00
520328210	4			Add subjects: Qur'an Hadist 3.54, Fiqh 4.96, Aqidah Akhlak 4.48; Islamic history 4.46, Arabic language 3.92, Tafsir 5.56, Hadist science 5.23
520328319	4		5.1	Grade for history; 5.75 for letters; and 5.65 for Germany
730230708	5			Danem from SMEA (senior high school level) had lesson accountancy cost 8.09 and finances cost 7.67
731531610	4	N	8.6	Grade for secretary lesson
731531612	2	K	6.2	Grade for sociology and anthropology joined together
340219801	3			On Danem, there's accountancy lesson with grade 4.00
320611103	5			Added the grade of "Danem" from elektronika komunikasi lesson are 3.80
520328207	8			Add subjects: Qur'an Hadist 7.45, Fiqh 6.30, Arabic language 6.50, Aqidah-Akhlak 6.5 , Tafsir science 7.70, Hadist science 7.00

Glossary

A–F

<i>adat</i>	Traditional law of a community.
<i>arisan</i>	A kind of group lottery, conducted at periodic meetings. Each member contributes a set amount of money, and the pool is given to the tenured member whose name is drawn at random.
Bahasa Indonesia	Standard national language of Indonesia.
<i>bidan</i>	Midwife, typically having a junior high school education and three years of midwifery training.
<i>bina keluarga balita</i>	child development program.
book	Major section of an IFLS questionnaire (e.g., book K).
BPS	Biro Pusat Statistik, Indonesia Central Bureau of Statistics.
CAFÉ	Computer-Assisted Field Editing, a system used for the first round of data entry in the field, using laptop computers and software that performed some range and consistency checks. Inconsistencies were resolved with interviewers, who were sent back to respondents if necessary.
CFS	IFLS Community-Facility Survey.
data file	File of related IFLS2 variables. For HHS data, usually linked with only one HHS questionnaire module.
<i>desa</i>	Rural township, village. Compare <i>kelurahan</i> .
DHS	Demographic and Health Surveys fielded in Indonesia in 1987, 1991, 1994, 1997.
<i>dukun</i>	Traditional birth attendant.
EA	Enumeration Area.
EBTANAS	Indonesian National Achievement Test, administered at the end of each school level (e.g., after grade 6 for students completing elementary school).

G–K

HH	Household.
HHID	Household identifier. In IFLS1 called CASE; in IFLS2 called HHID97.
HHS	IFLS Household Survey. IFLS1-HHS and IFLS2-HHS refer to the 1993 and 1997 waves, respectively.
IFLS	Indonesia Family Life Survey. IFLS1 and IFLS2 refer to the 1993 and 1997 waves, respectively.
IFLS1 re-release, IFLS1-RR (1999)	Revised version of IFLS1 data released in conjunction with IFLS2 and designed to facilitate use of the two waves of data together (e.g., contains IDs that merge with IFLS2 data). Compare <i>original IFLS1 release</i> .
interviewer check	Note in a questionnaire for the interviewer to check and record a previous response in order to follow the proper skip pattern.
<i>kangkung</i>	Leafy green vegetable, like spinach.
<i>kabupaten</i>	District, political unit between a province and a <i>kecamatan</i> (no analogous unit in U.S. usage).
<i>kartu sehat</i>	Card given to a (usually poor) household by a village/ municipal administrator that entitles household members to free health care at a public health center.
<i>kecamatan</i>	Subdistrict, political unit analogous to a U.S. county.
<i>kelurahan</i>	urban township (compare <i>desa</i>).
<i>klirik</i> , <i>klirik swasta</i> , <i>klirik umum</i>	Private health clinic.
<i>kotamadya</i>	Urban district; urban equivalent of <i>kabupaten</i> .
<i>kyai</i>	Muslim religious leader.

L–O

LDUI	Lembaga Demografi, Demographic Institute of the University of Indonesia.
Look Ups (LU)	Process of manually checking the paper questionnaire against a computer-generated set of error messages produced by various consistency checks. LU specialists had to provide a response to each error message; often they corrected the data.

L–O (cont.)

<i>madrasah</i>	Islamic school, generally offering both religious instruction and the same curriculum offered in public school.
<i>madya</i>	Describes a <i>posyandu</i> that offers basic services and covers less than 50% of the target population. Compare <i>pratama</i> , <i>purnama</i> , and <i>mandiri</i> .
<i>mandiri</i>	Describes a full-service <i>posyandu</i> that covers more than 50% of the target population. Compare <i>pratama</i> , <i>madya</i> , and <i>purnama</i> .
<i>mantri</i>	Paramedic.
<i>mas kawin</i>	Dowry – money or goods – given to a bride at the time of the wedding (if Muslim, given when vow is made before a Muslim leader or religious officer).
module	Topical subsection within an IFLS2 survey questionnaire <i>book</i> .
NCR pages	Treated paper that produced a duplicate copy with only one impression. NCR pages were used for parts of the questionnaire that required lists of facilities.
origin household	Household interviewed in IFLS1 that received the same ID in IFLS2 and contained at least one member of the IFLS1 household. Compare <i>split-off household</i> .
original IFLS1 release	Version of IFLS1 data released in 1995. If this version is used to merge IFLS1 and IFLS2 data, new IFLS1 IDs must be constructed. Compare <i>IFLS1 re-release</i> .
“other” responses	Responses that did not fit specified categories in the questionnaire.

P–R

panel respondent	Person who provided detailed individual-level data in IFLS1.
<i>peningset</i>	Gift of goods or money to the bride-to-be (or her family) from the groom-to-be (or his family) or to the groom-to-be (or his family) from the bride-to-be (or her family). Not considered dowry (see <i>mas kawin</i>).
<i>perawat</i>	Nurse.
<i>pasantren</i>	School of Koranic studies for children and young people, most of whom are boarders.
PID	Person identifier. In IFLS1 called PERSON; in IFLS2 called PID97.

P–R (cont).

PIDLINK	ID that links individual IFLS2 respondents to their data in IFLS1.
PKK	Family Welfare Group, the community women's organization.
PODES questionnaire	Questionnaire completed as part of a census of community infrastructure regularly administered by the <i>BPS</i> . Retained at village administrative offices and used as a data source for CFS book 2.
<i>posyandu</i>	Integrated health service post, a community activity staffed by village volunteers.
<i>praktek swasta, praktek umum</i>	Private doctor in general practice.
<i>pratama</i>	Describes a <i>posyandu</i> that offers limited or spotty service and covers less than 50% of the target population. Compare <i>madya</i> , <i>purnama</i> , and <i>mandiri</i> .
preprinted roster	List of names, ages, sexes copied from IFLS1 data to an IFLS2 instrument (especially AR and BA modules), to save time and to ensure the full accounting of all individuals listed in IFLS1.
province	Political unit analogous to a U.S. state.
<i>purnama</i>	Describes a <i>posyandu</i> that provides a service level midway between a <i>posyandu madya</i> and <i>posyandu mandiri</i> and covers more than 50% of the target population. Compare <i>pratama</i> , <i>madya</i> , and <i>mandiri</i> .
<i>puskesmas, puskesmas pembantu</i>	Community health center, community health subcenter (government clinics).
RT	Sub-neighborhood.
RW	Neighborhood.

S–Z

SAR	Service Availability Roster, CFS <i>book</i> .
SD	Elementary school (<i>sekolah dasar</i>).
SDI	Sampling form 1, used for preparing the facility sampling frame for the CFS.
SDII	Sampling form 2, used for drawing the final facility sample for the CFS.
<i>sinse</i>	Traditional practitioner.

S–Z (cont.)

SMP	Junior high school (<i>sekolah menengah pertama</i>). The same meaning is conveyed by SLTP (<i>sekolah lanjutan tingkat pertama</i>).
SMU	Senior high school (<i>sekolah menengah umum</i>). The same meaning is conveyed by SMA (<i>sekolah menengah atas</i>) and SLTA (<i>sekolah lanjutan tingkat atas</i>).
special codes	Codes of 5, 6, 7, 8, 9 or multiple digits beginning with 9. Special codes were entered by interviewer to indicate that numeric data are missing because response was out of range, questionable, or not applicable; or respondent refused to answer or didn't know.
split-off household	New household interviewed in IFLS2 because it contained a target respondent. Compare <i>origin household</i> .
SUSENAS 1993	1993 socioeconomic survey of 60,000 Indonesian households, whose sample was the basis for the IFLS sample.
system missing data	Data properly absent because of skip patterns in the questionnaire.
<i>tabib</i>	Traditional practitioner.
target respondent	IFLS1 household member selected for IFLS2 either because he/she had provided detailed individual-level information in IFLS1 (i.e., was a <i>panel respondent</i>) or had been age 26 or older in IFLS1.
tracking status	Code in preprinted household roster indicating whether an IFLS1 household member was a <i>target respondent</i> (= 1) or not (= 3).
<i>tukang pijat</i>	Traditional masseuse.
Version	A variable in every data file that indicates the date of that version of the data. This variable is useful in determining whether the latest version is being used.
<i>warung</i>	Small shop or stall, generally open-air, selling foodstuffs and sometimes prepared food.
