WORKING

PAPER

User's Guide for the Indonesia Family Life Survey, Wave 5

Volume 2

JOHN STRAUSS, FIRMAN WITOELAR AND BONDAN SIKOKI

WR-1143/2-NIA/NICHD

March 2016

This product is part of the RAND Labor and Population working paper series. RAND working papers are intended to share researchers' latest findings and to solicit informal peer review. They have been approved for circulation by RAND Labor and Population but have not been formally edited or peer reviewed. Unless otherwise indicated, working papers can be quoted and cited without permission of the author, provided the source is clearly referred to as a working paper. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. RAND® is a registered trademark.

RAND LABOR AND POPULATION

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". March, 2000. DRU-2238/1-NIA/NICHD.

For papers using IFLS3 (2000):

Strauss, J., K. Beegle, B. Sikoki, A. Dwiyanto, Y. Herawati and F. Witoelar. "The Third Wave of the Indonesia Family Life Survey (IFLS3): Overview and Field Report". March 2004. WR-144/1-NIA/NICHD.

For papers using IFLS4 (2007):

Strauss, J., F. Witoelar, B. Sikoki and AM Wattie. "The Fourth Wave of the Indonesia Family Life Survey (IFLS4): Overview and Field Report". April 2009. WR-675/1-NIA/NICHD.

For papers using IFLS5 (2014):

Strauss, J., F. Witoelar, and B. Sikoki. "The Fifth Wave of the Indonesia Family Life Survey (IFLS5): Overview and Field Report". March 2016. WR-1143/1-NIA/NICHD.

Preface

This document describes the design and implementation and provides a preview of some key results of the Indonesia Family Life Survey, with an emphasis on wave 5 (IFLS5). It is the second of seven volumes documenting IFLS5. The first volume describes the basic survey design and implementation.

The Indonesia Family Life Survey is a continuing longitudinal socioeconomic and health survey. It is based on a sample of households representing about 83% of the Indonesian population living in 13 of the nation's 26 provinces in 1993. 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 re-interview 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, was fielded on the full sample in 2000. IFLS4 was fielded in late 2007 and early 2008 on the same 1993 households and their splitoffs. IFLS5 was fielded in late 2014 and early 2015 on the same set of IFLS households and splitoffs: 16,204 households and 50,148 individuals were interviewed. Another 2,662 individuals who died since IFLS4 had exit interviews with a proxy who knew them well.

IFLS5 was a collaborative effort of RAND and Survey Meter. Funding for IFLS5 was provided by the National Institute on Aging (NIA), grant 2R01 AG026676-05, the National Institute for Child Health and Human Development (NICHD), grant 2R01 HD050764-05A1 and grants from the World Bank, Indonesia and GRM International, Australia from DFAT, the Department of Foreign Affairs and Trade, Government of Australia.

The IFLS5 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 5

WR-675/1-NIA/NICHD: *The Fifth Wave of the Indonesia Family Life Survey (IFLS5): Overview and Field Report*. Purpose, design, fieldwork, and response rates for the survey, with an emphasis on wave 5; comparisons to waves 1, 2, 3 and 4.

WR-675/2-NIA/NICHD: *User's Guide for the Indonesia Family Life Survey, Wave 5*. Descriptions of the IFLS file structure and data formats; guidelines for data use, with emphasis on using the wave 5 with the earlier waves 1, 2, 3 and 4.

WR-675/3-NIA/NICHD: *Household Survey Questionnaire for the Indonesia Family Life Survey, Wave 5.* English translation of the questionnaires used for the household and individual interviews.

WR-675/4-NIA/NICHD: *Community-Facility Survey Questionnaire for the Indonesia Family Life Survey, Wave 5.* English translation of the questionnaires used for interviews with community leaders and facility representatives.

WR-675/5-NIA/NICHD: *Household Survey Codebook for the Indonesia Family Life Survey, Wave 5.* Descriptions of all variables from the IFLS5 Household Survey and their locations in the data files.

WR-675/6-NIA/NICHD: *Community-Facility Survey Codebook for the Indonesia Family Life Survey, Wave 5.* Descriptions of all variables from the IFLS5 Community-Facility Survey and their locations in the data files.

WR-675/7-NIA/NICHD: *Dried Blood Spot User's Guide for the Indonesia Family Life Survey, Wave 5*. Descriptions of the dried blood spot field and assay procedures and data quality analysis.

Contents

Preface ii

Acknowledgments vi

1. Introduction 1

2. IFLS5 Data Elements Deriving from Prior Waves

HHS: Re-interviewing IFLS1 Households, and their Split-offs and Individuals 2
HHS: Pre-loaded Household Roster 3
HHS: "Intended" Respondents and Households 4
HHS: Obtaining Retrospective Information 6
HHS: Updating Kinship Information 6
Children 6
CFS: Re-interviewing IFLS Facilities and Communities 7

3. IFLS5 Data File Structure and Naming Conventions 8

Basic File Organization 8
Household Survey 8
Community-Facility Survey 8
Identifiers and Level of Observation 9
Household Survey 9
Community-Facility Survey 10
Question Numbers and Variable Names 12
Response Types 12
Missing Values 13
Unfolding Brackets 13
Special Codes and X Variables 15
TYPE Variables 15
Privacy Protected Information 16
Weights 16
IFLS5 longitudinal analysis household weights 16
IFLS5 longitudinal analysis person weights 19
IFLS5 cross-section analysis person weights 20
IFLS5 cross-section analysis household weights 21
Adaptive Number Series Test Data and Scores 21

4. Using IFLS5 Data with Data From Earlier Waves 23

Merging IFLS5 Data with Earlier Waves of IFLS for Households and Individuals 23
Data Availability for Households and Individuals: HTRACK and PTRACK 23
HTRACK14 24
PTRACK14 25
Merging IFLS1, 2, 3, 4 and 5 Data for Communities and Facilities 26

Appendix

A: Names of Data Files for the Household Survey 28
B: Names of Data Files for the Community-Facility Survey 33
C: Module-Specific Analytic Notes 38

Glossary 47

Tables 53

Acknowledgments

A survey of the magnitude of IFLS 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.

IFLS5 was directed by John Strauss (University of Southern California and RAND). Firman Witoelar (Survey Meter) and Bondan Sikoki (Survey Meter) were co-PIs. Sikoki and Witoelar were Field Directors of IFLS5.

Dr. John Giles of World Bank has helped to develop and oversee pretests and training of the community and facility (COMFAS) questionnaire modifications. He also contributed greatly in obtaining a grant from the research arm of World Bank to help fund collection of the community/facility data.

Dr. Peifeng (Perry) Hu of UCLA Medical School directed the laboratory analyses of the dried blood spots (DBS). Dr. Eileen Crimmins of USC provided critical assistance to the DBS part of IFLS and to the quality analysis of the DBS data. Dr. Elizabeth (Henny) Herningtyas of University of Gadah Mada directed the laboratory in Yogyakarta where the assays were done. Validation samples were collected at UCLA under the direction of Heather McCreath and further analyzed at Dr. Alan Potter's laboratory at University of Washington.

Many of our IFLS family colleagues have contributed substantially to the survey. Most of all, however, we are immensely grateful to Dr. Duncan Thomas and Dr. Elizabeth Frankenberg of Duke University and to Dr. James P. Smith of RAND, whose inputs continue to be invaluable and essential; and to Dr. John Phillips of the National Institute on Aging and Dr. Regina Bures of the National Institute of Child Health and Human Development, our project officers, whose continual encouragement and consultation has greatly benefitted this project.

Dr. John McArdle of USC oversaw the creation of the adaptive number series test that is new in IFLS5 (in Section COB of Book 3B). This involved extensive pretests in both Indonesia and Mexico (with the collaboration of MHAS under the direction of Rebeca Wong) and analysis of the pretest data.

Dr. Arthur Stone of USC Center for Economic and Social Research (CESR) and Dr. Jacqui Smith of the University of Michigan provided the hedonic well-being questions (Section PNA of book 3A) from their shorter version that they developed for the Health and Retirement Study (HRS).

Dr. Brent Roberts of University of Illinois and Dr. Angela Duckworth of University of Pennsylvania provided advice on the personality section (PSN in book 3B). Under their guidance we choose the BFI 15 questions as the ones we use.

Dr. Joan Broderick of USC CESR provided advice on the sleep quality and sleep disturbance questions in section TDR in book 3B. These questions were provided from PROMIS (Patient Recorded Outcomes Measurement Information System) staff.

Several people played critical staff roles in the project. Edy Purwanto was the Field Coordinator for the Household Survey, Naisruddin was Field Coordinator for the Community-Facility Survey, Iip Umar Rifaii was Field Coordinator for the Computer-Assisted Personal Interview (CAPI) and was responsible for the software development, and Roald Euller of RAND was Chief Project Programmer. Other key Survey Meter programmers were Nursuci Arnashanti and Amalia Rifana Widiastuti. Ika Rini was responsible for managing the dried blood spots when they got to Survey Meter from the field. This included logging in the cards, making sure that all labels were correct and person ids were on the samples, making a judgment and recording data on blood spot quality, and supervising their storage in the freezers and

monitoring freezer temperatures, Data quality checking at Survey Meter was headed by Dian Hestina, Lazima and Rini Kondesiana.

Strauss and Witoelar oversaw the construction of the sampling weights with the help of Yudo Wicaksono. Witoelar also oversaw the work to update geographic location codes using updated BPS location codes; as well as to update the IFLS "commid" community codes for the new areas in which split-off households were found in 2014. He also did most of the work in obtaining the tables and figures in the Field Report and the User's Guide. Wicaksono worked on the coding of the job-type and sector of work.

The IFLS5 public-use data files were produced with much painstaking work, by our Chief Programmer, Roald Euller. Euller also prepared the information used in the preloaded rosters and other preloaded files and the master household location files that were used in the field work.

The success of the survey is largely a reflection of the diligence, persistence and commitment to quality of the interviewers, supervisors, field coordinators and the support staff at our central headquarters in Yogyakarta. Their names are listed in the *Study Design*, Appendix A.

Finally, we thank all of our IFLS respondents both in households and communities for graciously agreeing to participate. Without their being willing to share their valuable time this survey could not have been successful.

Introduction

By the middle of the 1990s, Indonesia had enjoyed over three decades of remarkable social, economic, and demographic change. Per capita income had risen since the early 1960s, from around US$50 to more than US$1,100 in 1997. Massive improvements occurred in many dimensions of living standards of the Indonesian population. The poverty headcount measure as measured by the World Bank declined from over 40% in 1976 to just 18% in 1996. Infant mortality fell from 118 per thousand live births in 1970 to 46 in1997. Primary school enrollments rose from 75% in 1970 to universal enrollment in 1995 and secondary schooling rates from 13% to 55% over the same period. The total fertility rate fell from 5.6 in 1971 to 2.8 in 1997.

In the late 1990s the economic outlook began to change as Indonesia was gripped by the economic crisis that affected much of Asia. At the beginning of 1998 the rupiah collapsed and gross domestic product contracted by an estimated 13%. Afterwards, gross domestic product was flat in 1999. Between 2003 and 2014 GDP growth fluctuated between 5% and 6% per year and recovery ensued.

Different parts of the economy were affected quite differently by the 1998 crisis, for example the national accounts measure of personal consumption showed little decline, while gross domestic investment declined 35%. Across Indonesia there was considerable variation in the impacts of the crisis, as there had been of the earlier economic success. The different waves of the Indonesia Family Life Survey can be used to document changes before, during and 3,10 and 17 years after the economic crisis for the same communities, households and individuals.

The Indonesia Family Life Survey is designed to provide data for studying behaviors and outcomes. The survey contains a wealth of information collected at the individual and household levels, including multiple indicators of economic and non-economic well-being: consumption, income, assets, education, migration, labor market outcomes, marriage, fertility, contraceptive use, health status, use of health care and health insurance, relationships among co-resident and non- resident family members, processes underlying household decision-making, transfers among family members and participation in community activities.

In addition to individual- and household-level information, IFLS provides detailed information from the communities in which IFLS households are located and from the facilities that serve residents of those communities. These data cover aspects of the physical and social environment, infrastructure, employment opportunities, food prices, access to health and educational facilities, and the quality and prices of services available at those facilities.

By linking data from IFLS households to data from their communities, users can address many important questions regarding the impact of policies on the lives of the respondents, as well as document the effects of social, economic, and environmental change on the population.

IFLS is an ongoing longitudinal survey. The first wave, IFLS1, was conducted in 1993–1994. The survey sample represented about 83% of the Indonesian population living in 13 of the country's 26 provinces.[1] IFLS2 followed up with the same sample four years later, in 1997–1998. One year after IFLS2, a 25% subsample was surveyed to provide information about the impact of Indonesia's economic crisis. IFLS3 was fielded on the full sample in 2000, IFLS4 in 2007-2008 and IFLS5 in 2014-2015.

---

[1] Public-use files from IFLS1 are documented in six volumes under the series title *The 1993 Indonesian Family Life Survey*, DRU-1195/1–6-NICHD/AID, The RAND Corporation, December 1995. IFLS2 public use files are documented in seven volumes under the series *The Indonesia Family Life Survey*, DRU-2238/1-7-NIA/NICHD, RAND, 2000. IFLS3 public use files are documented in six volumes under the series *The Third Wave of the Indonesia Family Life Survey (IFLS3), WR-144/1-NIA/NICHD.* IFLS4 public use files are documented in the series *The Fourth Wave of the Indonesia Family Life Survey (IFLS4), WR-675/1-NIA/NICHD*.

2. IFLS5 Data Elements Deriving from Prior Waves

This section discusses elements of the IFLS5 data that derive from the earlier waves of IFLS. The bulk of the discussion applies to the household survey (*HHS*), [2] with the community-facility survey (*CFS*) covered at the end of the section.

Re-interviewing IFLS1 Households and Their Split-offs and Individuals

As explained in Sec. 2 of the *Overview and Field Report* (WR-/1-NIA/NICHD), IFLS5 attempted to re-interview all 7,224 households interviewed in IFLS1, plus all of the newly formed households (split-offs) that first appeared in IFLS2 through 4. The original IFLS1 households, plus the later wave split-offs, we call collectively our target households. For each of these *target households*, a roster was generated and preloaded into CAPI (see next section). It listed the household's IFLS ID from the last time it was found and the name, age, sex, birthdate, and relationship to the household head of all previous members of the household in the most recent interview. This preloaded roster included any person who had been listed as a household member in any prior wave. In addition, the preloaded information included each person's household status in 2007, whether books 3, 4 and 5 were completed in 2007 and the tracking status in 2014, which identifies whether the individual was a *target respondent* for tracking. It was target respondents who were to be tracked if they were not currently a resident of the household.

As in earlier waves of IFLS, interviewers were instructed to first return to the address where the household was last located. For each *HHID*, detailed address information was given in an "address book", on all past addresses lived in at the times the household was found in earlier waves. In addition, the address book had a list of all members ever found in the household, their names, sex, age and *PIDLINK* and their status (household member, moved, new member) for each prior wave. In addition we provided a "contact book" with information on all places of previous residence, places of past employment and schools where the children went, for each household. We also had, from previous waves, names and addresses of local contact persons. If the entire household was missing, the interviewers were instructed to look for all target members, if it was thought to still be in an IFLS province. If only individual members were not in residence as household members, those who were deemed to be target respondents were also tracked, if they were thought to still be in an IFLS province.

We continued the "first point of contact" rule, implemented in IFLS2, 2+, 3 and 4. At the point of first contact during the 2014 fieldwork with any IFLS household member, the household in which that person had resided in at the last interview was said to have been found. An interview was conducted using the same HHID as the last interview, with current information collected for everyone listed in the preloaded roster. As an example, suppose household 0930500 contained two members in 2007 but they divorced in 2009. If the member with PIDLINK 093050002 was located first, then that person is assigned the origin HIHD, 0930500. If the other member (PIDLINK 093050001, the previous household head) was later located, that person is identified as a (new) *split-off* household. In the vast majority of cases an *origin household* resided at the household's last known location and included most of the past members. As happened in prior waves, other scenarios also occurred, where the origin household resided:

- At a distant location from the last known dwelling but with the household intact
- At a different location with a few 1993 household members
- At the same dwelling but with very few of the 1993 household members.

---

[2] Italicized terms and acronyms are defined in the Glossary.

Application of the "first contact" rule for the target households[3] sometimes yielded some odd results. Some hypothetical examples are:

- In a 1993 household of 5 people, all had moved from the 1993 location by 2014. The 17-year-old son, not born in 1993, 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 HHID. 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 assign them the origin HHID and the son with his aunt's family the split-off household, but the rule dictated otherwise.

- A young man in a 1993 household marries in 2003 and the couple moved in with the wife's parents after marriage, which was tracked in 2007 and called a split-off household. The couple divorced by 2010, and both man and woman moved out of the woman's family's household. In 2014 the woman's family's household was found but with no target respondents for IFLS5, or their spouses or children. In that split-off household, no one would have been interviewed given the interview rules in place for IFLS5, but the household will show up in the database.

After contacting the household, the household roster (Book K, Module AR) is completed and all individuals are identified as being present or not (AR01a) and qualifying for an individual interview (AR01i). One way of spotting anomalies from the "first contact" rule 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). Alternatively, in split-off households, look for a large number of people who should not have been interviewed (AR01i = 3), either because they moved out (AR01a = 3) or because they did not meet the IFLS4 criteria for being interviewed. In using IFLS data generally, remember that not all individuals listed in the household roster for origin households were current members of the household in a particular wave. The household roster is meant to be a cumulative list of all household members found in that household in all waves of IFLS.

Another apparent anomaly is that for a small number of households, a household roster exists but includes no current members who were given individual books (AR01i=3 for all members). In these cases only part of the AR module of book K was filled out and the rest left missing because we were not interested in these particular households anymore. This occurred because there were no target respondents still alive in 2007 and residing in the household at the time of the IFLS5 interview.

Pre-loaded Household Roster

In certain modules, information collected in previous waves of IFLS was pre-loaded onto CAPI and used in interviews. The purpose was twofold: to ensure that information on particular households and individuals was updated and to save time during the interview.

The most important example of pre-loaded information (others are discussed later in this section) was the household roster. For every target household, a roster was generated that contained the following information for each IFLS1, 2, 2+, 3 or 4 household member:

---

[3] 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 preloaded 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.

Person Identifier (*PID*)

Person Identifier (PIDLINK)

Name

Sex

Age

Birthdate

Respondent's Household status last interview

Relation to the household head last interview

Tracking status in 2014 (whether the person was a target respondent)

Interview status in 2014 (whether the person was to get personal interview; relevant for split off households) in 2014

Panel status for books 3, 4 and 5 (whether the person gave detailed information in IFLS4 for books 3 or 4)

When a target household was found, the interviewer asked for updated information about each member on the list.

The pre-loaded roster was invaluable in making sure that IFLS5 collected at least some information about every 1993, 1997, 1998, 2000 and 2007 household member, as well as maintaining the person's id's within the household. When a target respondent had moved out of the household, his or her preloaded information was transferred onto a tracking form that was used to collect information about where the person had gone.

For new split-off households in 2014 we used a blank roster rather than preloaded roster as the base page in book K. All members of the new household were manually entered on the CAPI page. *PIDLINK*s (defined in Sec. 3) and panel status information were transferred from the tracking forms onto the base page for individuals who had been tracked from the target household to the new split-off household.

"Intended" Respondents and Households

In IFLS5, like earlier waves, we sought to re-interview all target households, plus new split-off households that contained at least one target respondent. Every household was administered at least one Book T to obtain contact information in case the household had moved, or to be used to find the household in the next wave. If the household was found, a knowledgeable household member was interviewed. If not, usually a neighbor was found. For obtaining household-level information, interviewers administered 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 or the spouse of a male 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 households that were found in IFLS5, we followed the practice of earlier waves and sought to interview all current members of an origin household. In split-off households, whether new split-offs in 2014 or split-offs from 1997, 1998, 2000 or 2007, we stayed with the practice of IFLS3 and 4 and interviewed any person who had been an original IFLS1 household member (regardless of whether they were the person being tracked), their spouse and biological children, if any. In actual fact

this did not make such a large difference because most of the non-target IFLS1 members were spouses or children of the target member.

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 household members from a previous wave, information in the preloaded roster indicated whether the person should answer books 3A and 3B or book 5, which was programmed in CAPI. In the field, interviewers sometimes encountered respondents who said they were younger than 15 but were listed as over 15 in Book K. In this case we overrode Book K, updated it and administered book 5. For new household members, age information was overridden if a parent insisted that the age of his or her child was different than what was reported in AR. If for example, a child was said to be age 16 in AR, but the parent later insisted the child was 13, then book 5 would be administered to the child instead of books 3a and 3b.

Information from parents about children and pregnancies were collected in both books 3B and 4. For women who were previous respondents, preloaded information indicated which of those books the woman should answer. If she had answered book 4 in 2007, she was asked to answer it again in 2014, so long as she was under 58, whereas book 4 was technically limited to ever-married women 15–49. So a woman who answered book 4 in 2007 and was under 50 years old then, also answered it again in 2014. That way we are able to get continuous marital and fertility histories of panel women from age 19 to 49. If a woman had not answered book 4 in 2007, perhaps because she was under 15 years old then, or never married, she was asked to answer it in 2014 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. As in prior waves, 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 adult respondents from whom we wanted information, we used a proxy book (Book Proxy), first introduced in IFLS2, 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 considerably less information than the “main” books. For example, we collected data about only two of a woman’s pregnancies in the last 4 years. 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 Book Proxy and corresponding main books in IFLS5. What was kept in Book Proxy is a little different than in IFLS2, 3 and 4, so it is worth the user’s while to compare Table 2.1 below with the corresponding table in the IFLS2, 3 and 4 User’s Guides (Frankenberg and Thomas, 2000; Strauss et al., 2004; Strauss et al., 2009). The questions are a subset of questions in the main books and so the questions have the same number in Book Proxy as they do in the main books. For example, TK25A1 contains information on last month’s earnings on the main job, both in book 3A and in Book Proxy. Different from what was done in past waves, IFLS5 data have the proxy information together with individual interviews for the same variable, say TK25A1. There is a separate variable, from the book cover of the proxy book, which indicates whether the respondent’s answers are by proxy. If so, then all answers in Books 3A, 3B and 4 will be by proxy. Thus to make full

use of the available individual-level information, the analyst no longer has to append data from Book Proxy to the related data from Books 3A, 3B, and 4. The data are already combined for IFLS5.

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 non-response for individuals and households, respectively for IFLS1, 2, 3, 4 and 5.[4]

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.

We followed the past practice in that respondents who had provided detailed information in IFLS4 (i.e., panel respondents) were not asked to provide full histories again in IFLS5. The criterion we used was that respondents who had answered books 3A, 3B or 4 in IFLS4 were considered panel respondents and in many cases only updated the information they had provided previously. For respondents who had not answered Books 3A, 3B, or 4 in IFLS4, we requested 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 preloaded 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.

IFLS5 generally collected less information about panel respondents than about new respondents. The questionnaires in IFLS5 were structured (1) to collect the same retrospective information for new respondents as had been collected in prior waves, and (2) for panel respondents, only to update the information collected in previous waves with information about what had happened since a particular point in time, mostly since the IFLS4 survey, but not always. To help prompt the respondent about the events for which we had data, preloaded information were sometimes made available to interviewers depending on the section. Therefore, to provide full retrospective information for IFLS5 panel respondents, the analyst must link data from all past waves.

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.

Updating Kinship Information

In past waves of IFLS certain respondents were asked very detailed information about their siblings and children. In IFLS5 we continued to ask detailed questions about biological and non-biological children, but for siblings we only ask about transfers at an aggregated level for all siblings. Rather than burdening respondents with the time-consuming task of re-listing the children in IFLS5, we preloaded rosters of children for interviewers to use.

Children

In IFLS5, we preloaded child rosters for panel respondents for module BA who had provided information on their children in IFLS4 in modules BA and/or CH and thus were expected to be eligible for the BA

---

[4] These files are described in more detail in Sec 4.

module in IFLS5. Rather than limiting the rosters to children not residing in the household in 2007, we listed all living children reported in 2007 in sections BA, CH, BX and AR.

IFLS5 respondents who did not provide child information in 2007 (so did not have a preloaded child roster), but were eligible to do so in 2014, completed a complete BA child roster. That group included men whose wife was no longer a household member, women who had answered book 3 or book 4 in 2007 but who had no children at that time, women who were not found in 2007 and new respondents.

More details about module BA appear in Appendix C.

Re-interviewing IFLS Facilities and Communities

Whereas a primary goal of the household survey was to re-interview households and individuals interviewed in previous waves, the community-facility survey aimed at describing the communities and available facilities for households and individuals interviewed in IFLS5. We sought to maintain comparability with instruments from the prior waves, but we were not explicitly trying to obtain high re-contact rates for facilities or specific respondents interviewed in communities or facilities in the past.

At the community level for all waves of IFLS, we sought interviews with two officers of the community: the head of the community, the *kepala desa* or *kepala kelurahan*, and the head of the local women's group, *PKK*. To the extent that there was continuity in the holders of those positions, the same individuals were interviewed in all waves. For community-level information, we have not attempted to determine whether particular respondents in 2014 were also respondents in earlier waves.

With respect to facilities, the same sample selection procedure was used in IFLS5 as in previous waves. To the extent that there was little turnover in the facilities available to respondents and that few facilities were available in a particular stratum to sample from, many of the facilities interviewed in 2007, 2000, 1997 or 1993 were interviewed again in 2014. To the extent that there was facility turnover or many facilities exist in a sampling frame, there may be low re-contact rates. This will be so for private health facilities, for example, because of the large number and turnover of that type.

To assist in matching facilities across waves, we assigned facilities which had been in prior waves, the same ID, the variable FCODE.[5] In the field, reassignment of the 1993, 1997, 2000 and 2007 IDs to a facility was accomplished with the Service Availability Roster (SAR). We preloaded this roster from IFLS4 for all community-facility survey teams. The preloaded SAR included a list of the names, addresses, and IDs of facilities mentioned in previous waves of IFLS as being available within the EA. Completing the SAR required (1) noting whether each facility on the preloaded list was still available in 2014 and (2) listing any facility newly available to community members since IFLS4 that was identified by either a household survey respondent or a community informant. In using the SAR to finalize the facility sampling list, the field supervisor assigned the 1993, 1997, 2000 or 2007 ID, FASCODE, to any facility noted as still being available in 2014.

[5] The exception is community health posts (*posyandu*). No community health post interviewed in IFLS5 has the same ID as its previous IFLS counterparts. That is because both the locations and volunteer staff changed over time, so determining whether an IFLS5 post was the same as in the past was effectively impossible. It is perhaps more appropriate to regard a community health post as an activity rather than a facility.

3. IFLS5 Data File Structure and Naming Conventions

This section describes the organization, naming conventions, and other distinctive features of IFLS5 data files to facilitate their use in analysis. The basic principles have not changed from prior wave releases. Additional information about the data files is provided in the survey questionnaires and codebooks. For analysts' convenience, each page of the household survey and community-facility survey questionnaires includes the names of data 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 household and community-facility data are available in SAS vxx? and Stata v13.0 formats.

Household Survey

The organization of IFLS5 follows closely that for prior waves. Household 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 at least 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.

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. In some cases the data file numbering sequence is out of order, where questions from previous IFLS waves have been dropped. For example, due to some changes in Book 5, Module DLA, we now have B5_DLA1, B5_DLA3 and B5_DLA4. Appendix A lists the name of each data file from the IFLS5 household survey, along with the associated level of observation and number of records.

Community-Facility Survey

Community-facility 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.

Data files are named by the questionnaire book and follow the same convention as names of household files.

For example, consider book 1, module A, data file BK1_A. The first page of the questionnaire 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 a different data file, BK1. Data 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 IFLS5 community-facility survey, 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 HHID14 uniquely identifies an observation. If the level of observation is the individual, both HHID14 and PID14 are required to uniquely identify a person, unless PIDLINK and AR01a are used.[6]

In IFLS5, HHID14 is a seven digit character variable whose digits carry the following meaning:

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 either 1, 2, 3, 4 or 5 depending on which wave the split-off first appeared. Split-offs from IFLS2 have their sixth digit equal to 1, while split-off households first appearing in IFLS2+ have a 2, split-offs from 2000, a 3, split-offs in 2007 have a 4 and new split-offs from 2014 a 5. The 7th digit indicates whether it is the first, second, or other split-off (some multiple split-offs occurred) for that wave.

In IFLS5, the person identifier PID14 is simply the line number of the person in the AR roster. It is possible that the PID number can be different for the same person, across waves if they reside in different households. Because of this PIDLINK is preferred way to link individuals across waves of IFLS.

When the level of observation is something other than the household or individual, it is usually because 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 Book 1, 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 HHID14 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 12 provider types for each of xxx households. Thus, there are 12 $\times$ xxx = yyy 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 the last visit in the last month by an individual to an outpatient provider. Not all individuals made visits, so some individuals appear only once, others never

[6] Within IFLS5 files, use HHID14 and PID14 to identify individuals. In the IFLS5 AR roster, variable PIDLINK does not uniquely identify individuals because individuals can be listed in more than one household roster. However, they are a current member of only one household, so PIDLINK together with AR01a=1, 5 or 11 can uniquely identify a household member.

appear. 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 HHID14, PID14, and RJTYPE.

Community-Facility Survey

Wherever possible, community-facility survey data are organized so that the level of observation within a data file is either the community or the facility. In a community-level data file, an observation can be uniquely identified with COMMID14. In a facility-level file, an observation can be uniquely identified with the variable FCODE14.

The first two digits of variable COMMID14 identify the province, and the remaining two digits indicate a sequence number within the province:

x x	x x
Province	Sequence

The following codes identify the 13 IFLS provinces[7]:

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	

COMMID14 are digits for the 312 communities (311 in IFLS5) that correspond to the 321 EAs, and for a common EA COMMID14 will be identical with COMMID00 and COMMID07.[8] For mover households, if they moved to a non-IFLS community, COMMID14 contains letters as well as digits, and is patterned after COMMID97, 00 and 07. In this case the first two digits still represent province, the third character represents district within province and the fourth sub-district within district. While for households within the original IFLS1 EAs, the level of COMMID14 is the EA-level (except for the 9 twin EAs), this is not true for movers outside of the original 321 IFLS1 EAs. For movers COMMID14 is generally at the sub-district level, as is COMMID00 and COMMID07.

For mover households in non-IFLS EAs, we now have Mini-CFS as a source of community data. As for IFLS3 and 4, we have created a separate community identifier for this module, MKID14. The structure of MKID14 is five characters, taking COMMID14 as its base and then adding one more character, that can be numeric or letter, that indicates the local area within the sub-district. For households living in one of the 321 IFLS EAs, MKID14 is just COMMID14 with a 0 as the fifth character. We did not want to use COMMID14 to identify area for Mini-CFS, because Mini-CFS was fielded at the EA-level, the local office

---

[7] We use the old BPS province definitions here to keep consistency with COMMID07. The location codes in SC update using the 2014 BPS codes, so provide province numbers for new provinces, carved out of West Java, for example. For mover households, the province code in COMMID14 are the old BPS province codes as of 1999, the same as for stayer households.

[8] Remember that 18 EAs (9 pairs) are so-called twin EAs, that are right next to each other and so arguably have the same conditions. These 9 pairs EAs are combined for the purpose of assigning a COMMID, so that there are only 312 COMMIDs, less the one that disappeared by 2014, so 311 in IFLS5.

of the *kepala desa* or *kepala kelurahan* being the source. Since there are sometimes multiple households in a sub-district, therefore having the same COMMID14, but possibly living in different EAs, it was necessary to have an identifier at the EA-level; MKID14 accomplishes this.

The first four digits of variable FCODE14 are the COMMID14 *of the place where the facility was first found*, the fifth digit indicates the facility type, and the last three digits indicate the facility type's sequence number within the community.

x x x x	x	x x x
COMMID14	Facility type	Sequence

The codes for facility type are the following:

0 = traditional health practitioner

1 = health center or subcenter (*puskesmas* or *puskesmas pembantu*)

2 = private practitioner (dokter *praktek* , *klinik swasta, klinik umum, bidan, bides, perawati, mantri*)

3 = private practitioner (*bidan, perawat, mantri,, this code is only for 1993 facilities*)

4 = community health post (*posyandu*)

5= community health post for the elderly (posyandu lancia)

6 = elementary school

7 = junior high school

8 = senior high school

9= hospitals

The codes of sequence shows in which wave the facility was found for the first time. The sequence < 70 shows facilities from IFLS1, 70 $\leq$ sequence < 100 from IFLS2, and 100 $\leq$ sequence < 249 from IFLS3 and 250 $\leq$ sequence < 500 from IFLS4 and 500 $\leq$ sequence from IFLS5.

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 now located, or the community for which the facility was interviewed, or the only IFLS community to which the facility provides services. To identify which facilities provide services to an IFLS community, analysts should use the Service Availability Roster (SAR).

Each SAR is a listing of all facilities, by type, that have served each COMMID since 1993. The SAR is organized by COMMID14 and FCODE14. As mentioned, some facilities serve several COMMIDs and so are listed in several SARs. Their FCODE14 will be the same in each of the SARs. The variable COMMID14 in the SAR file is the COMMID of the community for which the SAR is applicable, whereas, as discussed, the first four digits of FCODE14 are the COMMID of the location where the facility was first found.

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 COMMID14 or FCODE14 (if the data are from a facility questionnaire) and XXXTYPE for that data file. For the SAR, it is necessary to use both COMMID14 and FCODE14 to uniquely identify an observation because some facilities were shared by multiple communities, so an FCODE14 may appear more than once in the SAR.

Question Numbers and Variable Names

Most IFLS 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 IFLS5 questionnaire we tried to number the questions so as to preserve the correspondence with question numbers from earlier waves. If a question was added or changed in IFLS5, we typically added "a" or "b" to the question number rather than renumbering questions and destroying the correspondence. Since this had been done in IFLS2, 2+, 3 and 4, in IFLS5 one will sometimes see question numbers that have multiple letter extensions, such as XXXXaa or XXXXab, or XXXXAb.

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. We have tried to keep the response types identical across waves for the same question number or type.

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.

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. In IFLS, 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 CAPI 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), CAPI 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).

Unfolding Brackets

Beginning in IFLS4, to reduce non-response in questions reporting amounts of financial variables such as assets and income, we introduced the unfolding brackets questions. Respondents who refused to answer the question about the value of farm land owned by the household or answered don't know, would be asked a series of unfolding brackets questions. In IFLS5 every branch has three options, greater than, about equal to, or less than. In IFLS4 we had allowed only two options, greater than or equal to (an inequality) or less than. Adding the about equal to option follows practice in HRS. For example, first the respondent would be asked whether the value was more than Rp 20 million, about equal to Rp 20 million, or less than Rp 20 million. If the respondent answered that it was more than Rp 20 million, the respondent would be asked whether it was more than Rp 40 million, about equal to Rp 40 million, or less than Rp 40 million. If the answer was "less than Rp 20 million", the interviewer would then ask whether it was more than, about equal to, or less than Rp 10 million.

Based on the responses, the reported farm land value would be in one of these possible brackets: 0- Rp 10 million; Rp 10 million – Rp 20 million; Rp 20 million - Rp 40 million; Rp 40 million or more; or sometimes approximately equal to Rp 10 million or Rp 20 million or Rp 40 million. Although we still don't know the actual values of the farm land, these ranges of values will provide useful information that can be used in imputation.

In the example above, the values Rp 10 million, Rp 20 million, and Rp 40 million are called “break points”, and the value used in the first question of the unfolding brackets is called the “entry point”. In IFLS4, the values of the break points and entry points vary by the type of assets or income, but the entry point used was always the middle break point.

In IFLS5, to reduce the potential response bias that might arise by the choice of the entry point, the entry point was randomized by CAPI. Using the example above, one respondent who refused to answer the value of farm land may be asked whether it was more than, equal to, or less than Rp 20 million (the entry point randomly chosen was Rp 20 million). Then depending on the answer the interview will proceed as in the example above, either to the lower break point (Rp 10 million), or the higher break point (Rp 40 milion), unless the answer was approximately Rp 20 million, in which case the open brackets were ended. However, a different respondent may be asked with a different entry point: whether it was more than, equal to, or less than Rp 10 million. If the answer was “more than Rp 10 million”, then the next question will be whether it was more than, equal to, or less than Rp 20 million. If the answer was “more than Rp 20 million”, then the respondent will be asked again whether it was more than, equal to, or less than Rp 40 million. Figure below illustrates the possible flows of the questions.

Entry point Rp 20 million

11. < 20 million
    111. < 10 million
        112. ≈ 10 million
        113. > 10 million
        118. DON’T KNOW
12. ≈ 20 million
13. > 20 million
    131. < 40 million
    132. ≈ 40 million
    133. > 40 million
    138. DON’T KNOW
18. DON’T KNOW

Entry point Rp 10 million

21. < 10 million
22. ≈ 10 million
23. > 10 million
    231. < 20 million
    232. ≈ 20 million
    233. > 20 million
        2331. < 40 million
        2332. ≈ 40 million
        2333. > 40 million
        2338. DON’T KNOW
    238. DON’T KNOW
28. DON’T KNOW

The values of the data indicate which branch points the respondent took, for example 133 in the example above. The label indicates the value range of the last interval. For the example 133, the label would read "> 40 million". Or if the data point were 132 the label would read "approximately 40 million".

In addition randomizing the entry points, in IFLS5 we also expanded the use of unfolding brackets questions to many other financial variables throughout the household interviews, including income in section AR, all type of assets in section UT, HR, and NT, labor income and profits in section TK as well as values of assets lost in natural disaster (section ND).

Special Codes and X Variables

Many IFLS 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.

Privacy-Protected Information

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.

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 IFLS5 sampling weights that have been constructed for use with the household data. An overview of the weights from IFLS1, 2, 3, and 4 is provided in Table 3.1. The reader should consult the IFLS1, IFLS2, IFLS3 and IFLS4 User's Guides for details concerning those weights.

There are two types of weights for IFLS5 respondents. In constructing these we follow the overall procedures used to construct weights for earlier waves, with some alterations because of the inherent differences in having five waves instead of only three or four. The IFLS5 longitudinal analysis weights are intended to update the IFLS1 weights for attrition so that the IFLS5 panel sample (those IFLS5 households or individuals who were IFLS1 households or members in 1993, or their spouses or children), when weighted will be representative of the Indonesian population living in the 13 IFLS provinces in 1993. All respondents who were interviewed in IFLS5 but were not in an IFLS1 household roster are not assigned longitudinal weights; those will be missing in the data. We have also constructed longitudinal analysis weights for panel households and individuals who were in all five full waves of IFLS (IFLS1, 2, 3, 4 and 5). These weights are also intended to make this sub-sample of households or individuals representative of the 1993 population. For users who would rather not use inverse probability weights to correct for attrition (there are numerous assumptions required to properly use these weights), they can use the 1993 household or individual weights with the 2014 data, to get to 1993 population estimates that correct for the IFLS sample design, or they can create their own attrition correction factors.

The IFLS5 cross-section analysis weights are intended to correct both for sample attrition from 1993 to 2014, and then to correct for the fact that the IFLS1 sample design included over-sampling in urban areas and off Java. The cross-section weights are matched to the 2014 Indonesian population, again in the 13 IFLS provinces, in order to make the attrition-adjusted IFLS sample representative of the 2014 Indonesian population in those provinces. We also report cross-section weights that only correct for sample design, and not for attrition, just like the longitudinal weights.

IFLS5 longitudinal analysis household weights

Analyses of IFLS5 household data should use HWT14La (defined below) to obtain estimates that are weighted to reflect the Indonesian population in the 13 IFLS provinces in 1993. Panel analyses that use households in all four waves: IFLS1, 2, 3, 4 and 5 should use HWT_5_WAVES_L for the same end.

If all IFLS1 households were re-interviewed in IFLS5, the IFLS1 household weights and IFLS5 longitudinal analysis household weights would be identical. The IFLS5 longitudinal household weights therefore comprise two conceptually distinct components:

- Sample design effects that are embodied in the IFLS1 household weight, HWT93.
- An adjustment for household-level attrition between IFLS1 and IFLS5.

The IFLS1, 2, 3, 4, 5 longitudinal analysis household weight, HWT_5_WAVES_L, has the same two step design, except the attrition correction accounts for IFLS1 households that were not in IFLS2, 3, 4 and 5 and is estimated as a product of conditional probabilities. Fortunately, household-level re-contact rates in IFLS2, 3, 4 and 5 were very high (see the *Overview and Field Guide* for details). For users who prefer not to use any attrition corrections, they can use the weight HWT93 to correct for sample design effects.

Low attrition rates notwithstanding, adjusting for attrition is controversial because of assumptions that have to be made (see Wooldridge, 2010, for a good discussion). The main assumption is unconfoundedness, which means that variables that are unobserved (not used in the attrition logits) are uncorrelated with attrition.

For HWT14La we have followed the approach taken for IFLS3 and 4 and adopted the same simple model of between-wave (or jump-over) attrition, actually of being found. We first estimated a logit model of the probability that at least one member of an original IFLS1 household, or their spouse or children (even if they were first found in the dynasty household subsequent to IFLS1), was found in IFLS5, conditional on some basic household characteristics at the time of the first wave, IFLS1.[9] We use the same covariates that were used in deriving the IFLS2, 3 and 4 longitudinal weights. These include household size and composition in 1993, household location in 1993 and per capita household expenditure, also from 1993. Note that this specification ignores whether the household is found in 1997 (or 1998), 2000 or 2007. Estimates from these logit models are reported in Table 3.2. One can see that these models do well in explaining whether IFLS1 households are found in 2014.

We then computed the propensity score, the predicted probability the household was found, and inverted that probability to obtain an implied attrition adjustment for each household (inverted probability weight). That inverted probability becomes the essence of the attrition adjustment part of the weight. The attrition adjustments were then capped at the 99th percentile to prevent a single observation from receiving an inordinate weight. The product of the capped attrition adjustments and the IFLS1 household weight, HWT93, yield a household weight for each IFLS1 household that was found in IFLS5 that incorporates the original sampling design. We refer to this weight as ω_{HH1}.

The designs of IFLS2, 3, 4 and 5 called for following all *target respondents* (the definition of target varying some between waves as explained in Volume 1) who had moved out of the household by the time of the IFLS2, 3, 4 or 5 interview. Those target respondents who had moved generated *split-off* households and so a single IFLS1 household can spawn multiple IFLS2, IFLS3, IFLS4 and IFLS5 households. Indeed, as discussed elsewhere, multiple IFLS2 or 3 households sometimes merged together by IFLS3, 4 or 5. The split-off households complicate the construction of household weights. The IFLS5 household weights follow what was done for earlier waves' longitudinal weights and take this into account by distributing the estimated weight from the original IFLS1 household, ω_{HH1}, to the IFLS5 households spawned by that household. Specifically, assume κ IFLS1 household members, their spouses and children who may have joined the dynasty household later, were re-located in IFLS5; each of those IFLS5 respondents is assigned $(1/\kappa)$ of the

[9]Households in which all members of the IFLS1 households had died by 1997 or which combined with other IFLS households are treated as found in these calculations.

weight ω_{HH1} associated with their origin household. Taking the sum of these individual-assigned weights in the households in which they were found in 2014, yields the IFLS5 longitudinal analysis household weight. New household members since IFLS1, except for spouses and children of original members, thus do not contribute anything to the longitudinal household weight. The same procedure is used to derive the IFLS1, 2, 3, 4, 5 longitudinal analysis household weight (HWT_5_WAVES_L).

Note that this procedure is slightly different than that used for prior waves. In the past we only used IFLS1 original members to estimate the logits and distribute the dynasty household weights, but now that the sample is aging and many original members have died, we have adjusted this by adding spouses and children of IFLS1 members, whether or not they were IFLS1 members.

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 2/3 ω_{HH1} and the split-off household is assigned a weight of 1/3 ω_{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 IFLS5 household. In those instances, the calculation of the IFLS5 longitudinal analysis household weight follows the same principle and is the sum of individual-assigned weights based on the IFLS5 respondents' origin households in IFLS1.

To calculate the weight for HWT_5_WAVES_L we want the probability that a household was observed in all five waves. To estimate the this probability we could proceed in different ways. One approach would be to simply estimate a logit regression of the probability that an IFLS1 household was found in all five waves based on 1993 characteristics. It turns out that a better approach is to calculate conditional probabilities and use the product (Robins, Rotnitsky and Zhou, 1995; Wooldridge, 2010). Specifically, we estimate a logit for the probability that an IFLS1 household was found in IFLS2, then another logit for the conditional probability that an IFLS1 household found in IFLS2 was also found in IFLS3, likewise for the conditional probability that an IFLS3 household was found in IFLS4 and finally the same for the conditional probability a household found in IFLS4 was found in IFLS5. These conditional probabilities are estimated using covariates in the base period. So for the logit of being found in 1997, the sample is all of the IFLS1 households. The covariates in this logit are the same household variables with 1993 values described above, used to estimate the logit for HWT14La. For the logit regression for being found in 2014, conditional on being found in 2007, the sample is the set of dynastic households who were found in IFLS4. Covariates are the same as the logit of being found in 1997, except that the values are for 2007, the base period for this conditional probability.[10] Once we have the predicted conditional probabilities for each dynastic household that was found in all four waves, we multiply them together to estimate the unconditional probability of being found in all four waves. For households that do not appear in all four waves, this is set to missing.

These unconditional probabilities are then inverted, capped at the 99th percentile and multiplied by HWT93 to get the weights for the dynastic households. Finally, the same procedure as used for HWT14La is used to distribute the dynasty household weight to the component households in 2014.

To use these attrition corrections, we need to assume that unconfoundedness holds; that is that attrition depends only on the observed covariates in our logit equations. This is a strong assumption, and so there

[10] We use household size weighted averages of the sub-household variables for a reference dynastic household. For dummy variables such as province of residence or urban/rural location, we create new, combination dummy variables, such as a household has parts in both Jakarta and South Sulawesi provinces.

will be some users who are reluctant to use this part of the weights. As noted above, for the longitudinal weight HWT14a, they can instead use HWT93, which will correct for the sampling design.[11]

IFLS5 longitudinal analysis person weights

The IFLS5 longitudinal analysis person weights follow a similar approach. A longitudinal roster weight was first created by estimating a logit model of being found in 2014 for all individuals in the IFLS1 household rosters;[12] the model excludes all new entrants in IFLS2, 2+, 3 and 4. The covariates used are the same used first in calculating the weights for IFLS2. The inverse of the predicted probability yields the attrition adjustments. Estimates from the logit models are reported in Table 3.3. The covariates are the same as used in constructing the IFLS3 and 4 longitudinal person weights and are similar to those used for predicting households, but include a few more, felt appropriate for individuals.

The individual-specific attrition adjustments were also capped at the 99th percentile and multiplied by the IFLS1 household weight, HWT93, to take into account sample design effects;[13] the result is PWT14La. This IFLS4 longitudinal analysis person weight variable is recorded in PTRACK. PWT14La is not defined for any individuals in IFLS5 who were not listed in an IFLS1 household roster. Estimates that are weighted with one of these variables should correspond with the 1993 Indonesian population in the 13 IFLS provinces. Like the household longitudinal weights, if the user only wants to weight based on sample design effects, they should use HWT93.

A similar procedure as for the household weights was used to construct the longitudinal weights for being in all five full waves. For this purpose we need to consider another issue, that only a subset of IFLS1 roster individuals were chosen to be interviewed with individual books, so-called IFLS1 respondents. Most users will use information from individual books, hence the longitudinal weight we construct is for being a respondent in IFLS1 and in the IFLS2, 3, 4 and 5 waves. Note that this differs from our treatment of longitudinal weights for 2014 respondents, because in that case we construct weights, whether or not the person was a respondent in 1993, just that they were in the 1993 household.

For our purpose here, we take as our initial sample, those IFLS1 members who got individual books, and estimate a logit model for the conditional probability of these IFLS1 respondents being found in IFLS2. We also estimate logit regressions for the probability of being found in IFLS3 conditional on being found in IFLS2, for the probability of being found in IFLS4 conditional on being found in IFLS3 and for the conditional probability of being found in IFLS5 conditional on being found in IFLS4. These conditional propensity scores then get multiplied for those individuals in all four waves to arrive at the unconditional probabilities. As we do for our other weights, we then invert these and cap the inverted weights at the 99th percentile. Finally, we multiply the inverted, capped attrition adjustments by the IFLS1 individual weight, PWT93IN. PWT93IN adjusts both for the within household sampling in IFLS1, as well as uses the IFLS1 household

---

[11] This assumes that the unit of analysis is the dynastic household. If the observation is the sub-household, then HWT93 should be apportioned to the different sub-households, presumably by the fraction of the original IFLS1 household members who appear in the particular sub-household.

[12] An individual is considered found if the respondent was found in an IFLS4 household or is known to have died between the waves.

[13] Again, users who prefer not to use attrition corrected weights can simply use HWT93 to correct for sample design. Alternatively PWT93 can be used. The difference is that PWT also rakes into age and sex groups, not just province and urban-rural. It is the latter that makes the big difference in the IFLS sample design.

weight in order to make estimates representative of the underlying 1993 population. Our weight is named PWT_5_WAVES_L and is found in PTRACK.

The same procedure was followed to construct longitudinal analysis person weights for use with the health measures in Book US and a separate longitudinal weight was constructed for the DBS data. In IFLS1, a sub-sample of respondents were weighed and measured. In IFLS5, we sought to conduct physical health assessments on all respondents. Analyses using IFLS1, 2, 3, 4 and 5 measurements that want to be representative of the 1993 Indonesian population should use the weight PWT_5_WAVES_USL. This is based on a logit regression of all persons in the IFLS1 sample who were eligible to have US measurements (and thus have a positive and non-missing PWT93US from IFLS1) and estimates the joint probability that they had measurements taken in IFLS1, 2, 3, 4 and 5, using the same method of estimating logits for the conditional probabilities (see Table 3.4). For those panel members who did get health measurements taken in IFLS1, 2, 3, 4 and 5, the resultant predicted probabilities are inverted, capped and multiplied by the IFLS1 individual weight, PWT93US. The latter weight captures both the within household sampling in IFLS1 to choose who got measured, as well as the household sampling, to derive estimates representative of the 1993 population.

For the DBS data, the longitudinal weights can be used with the C-Reactive Protein data. The HbA1c data are new in IFLS5, so only cross-section weights are relevant. Since the DBS assays were done on a random subset of DBS samples and because DBS data was taken on a random sub-sample of IFLS4 extended households, the DBS weights are estimated a bit differently. We did not start assaying DBS until wave 4, another difference. For the longitudinal weight, PWT14_DBS_L, we construct a series of conditional probabilities and multiply them. First we start with any 1993 household member, whether or not they got individual books. We estimate the conditional probability they were found in 2007, using a logit model similar to that described above. We then apply sampling weights from 2007 which determine which extended households were to provide DBS samples. Next we apply a conditional probability model of whether the respondent actually gave DBS samples conditional on being in selected extended households; there were just over 19,000 respondents who gave DBS samples in 2007. We then estimate the conditional probability that the DBS samples were actually assayed, 9,945; not all were for cost and time reasons, but the ones that were chosen were randomly chosen, so this is not a logit, but the actual probabilities (see the C-Reactive Protein User Guide; Hu et al., 2013). Finally we estimate the conditional probability that a respondent who had a CRP assay performed in 2007 also did in 2014. Then we cap at the 99th percentile after inverting the product of the conditional probabilities.

IFLS5 cross-section analysis person weights

While IFLS is a longitudinal survey, there will be some analyses that treat IFLS5 as a cross-section. We have attempted to construct weights so that estimates based on IFLS5 will be representative of the Indonesian population living in the 13 IFLS provinces in 2014.

We have followed a procedure that parallels the approach taken to construct cross-section weights for IFLS2, 3 and 4. We rake the IFLS5 sample to an external sample, the 2014 wave of the SUSENAS in the 13 IFLS provinces, after having made adjustments for sample attrition from 1993 to 2014.

The attrition adjustments are made separately for IFLS1 household members and members who came into the household after IFLS1. We first drop households that are dead by 2014. For IFLS1 members we calculate the probability of being in IFLS5 given they were in IFLS1. We follow similar procedures to attrition adjustments for the longitudinal weights. We run a logit of being found in IFLS5 given they were in an IFLS1 household, using the same covariates discussed above, with 1993 values. We then invert the propensity scores and cap them as we do for the longitudinal weights.

For non-IFLS1 members who are in an IFLS5 household (whether or not they received an individual interview), we compute the probability of being in IFLS5 conditional on the household being in IFLS1 and being found in IFLS5, times the probability of the household being found in IFLS5, conditional on being in IFLS1. The latter probability we can get using the logits we use for the longitudinal household weight (using jumpover probabilities), distributing these propensities from the dynasty IFLS1 household to the component households in the same fashion as described above. The conditional probability of the non-IFLS1 person being found in IFLS5 given the household was found in IFLS5 and in IFLS1, is derived from a logit using the sample of all ever-members of an IFLS5 household, excluding IFLS1 members and members who died. Then the dependent variable is set to 1 if they are found alive in IFLS5 and the covariates include the same household level covariates used for the longitudinal weights with values from 1993 and individual variables dated the first time we saw this person in the IFLS household.

For the raking, all individuals listed as being present in the IFLS5 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). The IFLS5 cross-section analysis person weights are the ratio of the 2014 SUSENAS proportion to the IFLS5 proportion in each cell. These ratios of cell proportions have been re-weighted using the capped, inverted probability attrition adjustments calculated from the individual-specific logistic regressions in Table 3.3 The resulting weight is called PWT14Xa and is included in PTRACK. Estimates that use these weights should be representative of the Indonesian population in 2014 in the 13 IFLS provinces. As for the household cross-section weights, we also report the weights without attrition corrections, PWT14X_.

Similar weights have been constructed for use with the health assessments. PWT14USXa was constructed by raking IFLS5 for persons who had US measurements, to the 2014 SUSENAS, first taking into account attrition from 1993 to 2014 (from the IFLS1 roster to who was measured in IFLS5). Similarly, PWT14USX_ constructs the US weight without attrition adjustments.

We construct similar weights for the DBS sample, PWT14DBSXa was constructed by raking IFLS5 for persons who had DBS measurements, to the 2014 SUSENAS, after taking into account longitudinal attrition. For this we start from the 2007 longitudinal conditional probabilities that are described above, and then rake to the 2014 SUSENAS.

IFLS5 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 IFLS5 sample have been stratified by province and urban-rural sector. For each cell, the ratio of the proportion of households in the 2014 SUSENAS sample (in IFLS provinces) to the IFLS5 sample proportion, multiplied by the attrition-weight provides the IFLS5 cross-section analysis household weight, HWT14Xa. A second weight, HWT14X_, does not use the attrition correction. Estimates that are weighted with HWT14Xa should be representative of all households living in the IFLS provinces in Indonesia in 2014.

Adaptive Number Series Test and Scoring

The new adaptive number series test was added to IFLS5. It is patterned on the HRS adaptive number series test. Some of the questions overlap, while some are different, taken to be easier, more suitable for a low-income, low-education environment such as Indonesia. These questions were extensively pre-tested in both Indonesia and Mexico and a summary of the pretest, written by Dr. John McArdle and Dr. John Prindle is available on the IFLS web page to download. These questions are taken from the Woodcock-Johnson battery, same as used by HRS. The test is a measure of fluid intelligence, and can be compared to the Raven's test also used in IFLS.

The test is an adaptive test of six questions out of fifteen. The first three questions are taken by all respondents. Depending on how many of the three a respondent answers correctly, they will go to a second block of 3 questions that vary by difficulty. Within each block the three questions vary by difficulty. A total score for both sets of questions is then calculated, W-score, along with a standard error of the W-score. The W-scores are calculated from the table below, which was supplied by Dr. Richard Woodcock.

Starting Block refers to the initial three questions answered by all respondents. If a respondent got a 0 in the baseline three questions he/she went to block 1, if 1 correct block 2, etc. The cobxx_b variables are binary indicators if the answer to that question is correct or incorrect. The sum of those variables within a block is taken and then the table below used to obtain the W-score and standard error.

		Starting Block							
	# correct	0		1		2		3	
		W-Ability	SEM	W-Ability	SEM	W-Ability	SEM	W-Ability	SEM
Second Block	0	299	18	431	19	516	11	567	42
	1	322	20	467	20	527	9	604	11
	2	360	21	494	13	535	9	614	9
	3	396	19	510	12	545	10	625	11
	4							635	15

4. Using IFLS5 Data With Data From Earlier Waves

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

Merging IFLS5 Data with Earlier Waves of IFLS for Households and Individuals

The easiest method for merging household-level information is to use the variables HHID93, HHID97, HHID00, HHID07 together with HHID14. These are compatible in their construction and so one can safely merge at the household-level using these, after renaming them with the same name.[14]—Of course, not all households will merge. Some IFLS1 households were not re-interviewed in IFLS3 (or 2). And households that were new in IFLS5 will not have data in IFLS1, 2, 3 or 4.

To merge individual-level information across waves, use PIDLINK, which is available in IFLS1-RR, IFLS2, IFLS3, IFLS4 and IFLS5.

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]

The first 7 digits of PIDLINK indicate the household id where the person was first found. Do not merge across waves based on HHID14 and PID14, as you would within a wave. As an example, suppose that in IFLS1 the head's PERSON number was 01, his wife's number was 02, and their son's number was 03. By IFLS5 assume that all three members reside in different households. Assume that in IFLS5 the wife was contacted before the husband, who was contacted before the son. The range of identifiers for these individuals would be as follows:

	HHID93	PID93	PIDLINK	HHID14	PID14
Husband	1250100	01	125010001	1250141	01 (in split-off household)
Wife	1250100	02	125010002	1250100	02 (same as 93—still in origin)
Son	1250100	03	125010003	1250142	01 (in split-off household)

As we can see, combinations of HHID14 and PID14 may well not correspond to HHID93 and PID93, so one cannot match across waves on these variables. Thus PIDLINK is needed. It is the case that some PIDLINKs appear in two or more IFLS5 household rosters, because the rosters are cumulative from 1993. This means that PIDLINK by itself has nothing to do with which household in which the person was found in 2007. For the household(s) in which the person was not found in 2007, the Book K roster has AR01a = 3 (moved out of household), whereas AR01a=1, 2, 5 or 11 for the (one) household in which the person was found and interviewed. To avoid duplicate PIDLINKs, drop AR records where AR01a = 3. Also,

[14] This assumes that the re-released version of IFLS1 data files are being used.

PTRACK can be used to find the household that each person was found in, for each wave they were found.

Data Availability for Households and Individuals: HTRACK and PTRACK

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

HTRACK14

HTRACK14 contains a record for every household that was interviewed in IFLS1, 2, 2+, 3, 4 or 5. There are 13,536 household-level records in HTRACK14, one record for each of the 7,224 households that were interviewed in IFLS1 and one record for each of the additional 6,312 split-off households that were added in IFLS2, 2+, 3, 4 and 5 (2,425 splitoffs being new in IFLS5). HTRACK14 provides information on whether the household was interviewed in each wave (RESULT93, RESULT97, RESULT98, RESULT00, RESULT07, RESULT14) and, if so, whether data from books K, 1 and 2 are available. Codes for the result variables are:

1 = Interview conducted
2 = Joined other IFLS household
3 = All household members died
4 = Refused interview
5 = Not found
9= Missing[15]

HTRACK14 also provides information on the household's location in 1993, 1997, 1998, 2000, 2007 and 2014, if it was found. For 1993, five 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 (in the IFLS2 data) those used by BPS in 1999 (in the IFLS3 data), those used by BPS in 2007 and in 2014.[16] For 1997 locations, four sets of codes are given: those based on 1998 BPS codes, those based on 1999 codes, those based on 2007 codes and those based on 2014 codes. For 2000 locations we also provide three sets of codes: 1999, 2007 and 2014. For 2007 locations we give two sets of BPS codes, from 2007 and 2014. Finally for 2014 locations we provide 2014 SC codes. We use the 2014 BPS codes as the main set, and these are used consistently throughout IFLS5 (for example in module SC of books T and K). Note that using the 2014 codes is more difficult because two new provinces have been created from IFLS provinces in the 1993 codes: West Java was split into two as was South Sumatra.

For households that were interviewed in IFLS5, variable MOVER14 identifies whether the household moved between the last time it was interviewed (which could be 2007, 2000, 1998, 1997 or 1993).. MOVER14 takes the following values:

0 = Did not move
1 = Moved within same village/municipality
2 = Moved within same *kecamatan*

---

[15] Households with all members having died by IFLS2, 2+ or 3 have result07 set equal to 9, missing.

[16] 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. In general the BPS codes come out in June or July of a given year. These are the codes that get used in the SUSENAS fielded in February of the following year. So the 2014 BPS codes are the ones used in the 2015 SUSENAS (as well as SAKERNAS and other household surveys). 2014 codes and names for provinces, districts and sub-districts are contained in Table 4.1.

3 = Moved within same *kabupaten*
4 = Moved within same province
5 = Moved within other IFLS province

MOVER14 is non-missing not only for origin households interviewed in 2014, but also for split-off households interviewed in IFLS2, 2+, 3 or 4. In addition, we calculate MOVER14 for new split-off households in IFLS5. Because each split-off household contains at least one person who was tracked from an IFLS household (which could have been an origin household or could have been a split-off), we have calculated MOVER14 for split-off households on the basis of the household's 2014 location relative to the last known location of the household from which the tracked person came.

In addition to the BPS location codes, HTRACK14 contains COMMID93, COMMID97, COMMID00, COMMID07 and COMMID14, which can be used to link households to the IFLS community-level data. COMMID, described in detail above, is a four digit/character code. The first two digits represent the province, the third the district within province and the fourth the sub-district within district. All households found in a particular wave have non-missing COMMID for that wave, even if they are movers. COMMIDs for movers tend to have letters as their third or fourth characters. COMMID is defined at the level of the sub-district for mover households. For stayers, COMMID is defined at the enumeration area, except for the nine twin EAs, for whom their EAs are combined into one COMMID. This will allow users to estimate models with COMMID fixed effects, for example.

However for movers outside of the IFLS EAs, COMMID14 is not of help in linking to community data. MKID14 must be used instead to link to Mini-CFS, because it is defined at the EA-level for movers, not at the sub-district level. MKID14 is a five digit or character code, which contains COMMID14 as the first four characters, followed by a number or letter signifying EA within sub-district. Never-movers have a 0 as the fifth digit, whereas movers have a non-zero number or a letter. MKID14 should be used to match mover households to their Mini-CFS data files.

HTRACK14 also contains household weight variables, discussed above, for IFLS1, 2, 3, 4 and 5, both cross-section and longitudinal weights.

PTRACK14

PTRACK14 contains a record for every person who has ever appeared in an IFLS household roster. PTRACK14 contains 50,579 records, one for each of the 33,081 individuals listed in a 1993 household roster, and one for each of the additional 17,498 household members who have joined origin and split-off households since 1993 (43,649 in 2007, so 6,930 new individuals in IFLS5).

Within PTRACK14, each observation is identified by PIDLINK. PTRACK14 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. PTRACK14 indicates in which household each person who was ever an IFLS household member was found, in each wave, HHID93, 97, 00, 07, 14; plus their person IDs (PID) with the household in each wave. Further MEMBER93, 97, 00, 07, 14 indicates whether the person was indeed found in that wave. Individuals who moved out of the 1993 origin household and were interviewed in a new household will have different HHID and PIDs across waves. Individuals who were new household members in 2014 will have missing HHID and PID for 1993, 1997, 2000 and 2007.

We calculate our best guess of each person's age at each wave: AGE93, 97, 00, 07, 14. We also report our best guess of the person's date of birth. AGE93, AGE97, AGE00 and AGE07 are taken from the IFLS2, 3 and 4 PTRACKs and so represent the best guess age in 1993, 1997, 2000 and 2007 using information available in IFLS4, 3, 2 or 1. AGE14 is our best guess age in 2014 based on IFLS5 information. These will not necessarily be consistent across waves, although the algorithm that generates them is essentially the same. In theory respondents interviewed in IFLS4 should have been seven or eight years older in 2014/15, 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 and ages across waves do not always match for a respondent, and there may even be discrepancies between books within a wave. In addition to age and date of birth, we report our best guess of the person's sex based on IFLS5 data. For all but a few respondents, the reported sex matches across waves. The PTRACK14 file provides our best guess for sex in an attempt to resolve discrepancies.

PTRACK14 also reports information on marital status at each wave and the survey books for which data are available from each wave. Such information allows the analyst to calculate the number of observations in IFLS1, 2, 3 and 4 and the number of panel observations for the various survey books.

PTRACK14 does not provide information on individuals' locations. At the household level, that information is in HTRACK14. For individuals who were new household members in 2014 (AR01a_07 = 5), the location information in HTRACK14 for 1993, 1997, 2000 and 2007 is not necessarily the location where the new individual resided in those years. The individual's household of residence from past waves, in PTRACK14, can be used together with the location information in HTRACK14 to obtain past location, so long as the person was present in an IFLS household in that particular wave. Otherwise, to ascertain where a new household member lived in the past, data from module MG in book 3A should be used.

PTRACK14 also contains individual weights variables, described above, from IFLS1, 2, 3, 4 and 5.

Merging IFLS Data for Communities and Facilities

The IFLS database can be used as a panel of communities and facilities. In IFLS1, 2, 3, 4 and 5 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 changes in community characteristics over time.

In IFLS5, IFLS4, IFLS3, IFLS2, and IFLS1-RR data files, variable COMMID identifies the IFLS communities, with an extension of 93, 97, 00, 07 or 14 to indicate the source year. In IFLS1, communities were identified by the variable EA. The COMMID variables should now be used to link households with communities for non-mover households or households that moved to an IFLS EA. For movers to a non-IFLS EA use MKID14 to link household data files to the Mini-CFS data file.[17]

In IFLS1, 2, 3, 4 and 5, 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 IFLS1-RR and IFLS2, facilities are identified by the seven-digit character variable FCODE. In IFLS5, like prior waves, facilities are identified by the eight digit character variable FCODE14 (see Section 3 for a fuller description of FCODE14).

FCODE in IFLS1-RR and IFLS2 is a seven character code with the same structure as FCODE14 for the first 5 characters, and only 2 characters for facility number. Thus to convert the earlier FCODE to FCODE14 insert a 0 after the 5th character (for strata).

---

[17] In 1993, all IFLS households lived in one of 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 contained the three-digit EA code since it identified the community from which they moved, but it did not identify the community of their current residence. The same will be true for IFLS5. Analysts should not merge households with community data based on EA embedded in HHID, for that would link movers to communities in which they no longer live.

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, 3 and 4, all types of private practitioners received the same questionnaire and data are stored in the same files. To combine IFLS1 data from private practitioners with data from later waves, the analyst should first combine the IFLS1 doctor/clinic data with the IFLS1 nurse/paramedic/midwife data. In IFLS1 and IFLS2, all of school levels were administered in different questionnaires and stored in different files. In IFLS3, 4 and 5, all level of schools received the same questionnaire and data are stored in the same file, but the cover indicates which level. To combine IFLS1 and 2 schools data with data from IFLS3, 4 or 5, the analyst should first combine all of the schools level data of IFLS1 and IFLS2.

Appendix A:
Names of Data Files for the Household Survey

File Name	Contents	Level of Observation	Variable(s) that Identify the Unique Observation	No. Records
BT_COV	BK T Cover (Tracking Book)	Household	HHID14	18,277
BK_COV	BK K Cover (Control Book)	Household	HHID14	15,749
BK_SC1	BK K Location and sampling	Household	HHID14	15,921
BK_AR0	BK K Household size	Household	HHID14	15,921
BK_AR2	BK K Household roster	Individual	HHID14, PID14	89,384
BK_KRK	BK K Household characteristics	Household	HHID14	15,921
BK_TIME	BK K Date and time of interview	Household interview time	HHID14, TIME_OCC	
B1_COV	BK 1 Cover (HH Economy)	Household	HHID14	15,349
B1_KS0	BK 1 Consumption (1)-Misc	Household	HHID14	15,144
B1_KS1	BK 1 Consumption (2)-Food	Food expenditure item	HHID14, KS1TYPE	560,328
B1_KS2	BK 1 Consumption (3)-Non food monthly	Non food expenditure item	HHID14, KS2TYPE	181,692
B1_KS3	BK 1 Consumption (4)-Non food annually	Non food expenditure item	HHID, KS3TYPE	105,973
B1_KS4	BK 1 Consumption (5)-Prices	Food item	HHID14 KS4TYPE	151,350
B1_KSR1	BK 1 Assistance (1)- Cash Transfer program	Type of assistance	HHID14, KSR3TYPE	45,402
B1_KSR2	BK 1 Assistance (2)-misc	Household	HHID14	15,134
B1_KSR3	BK1 Assistance (3)-Food Assistance	Type of assistance	HHID14, KSR2TYPE	60,532
B1_KSR4	BK1 Assistance (4)-Disaster Aid	Type of assistance	HHID14, KSR5TYPE	30,266
B1_PP	BK 1 Health facilities	Facility	HHID14, PPTYPE	181,584
B1_TIME	BK 1 Date and time of Interview	Household	HHID14, TIME_OCC	23,369
B2_COV	BK 2 Cover- misc	Household	HHID14	15,351
B2_KR	BK 2 Housing characteristics	Household	HHID14	15,185
B2_UT1	BK 2 Farm business (1)-land, income	Household	HHID14	15,185
B2_UT2	BK 2 Farm business (2)-paddy harvest-grid	Paddy harvest	HHID14	2,447
B2_UT3	BK 2 Farm business (3)- crop hardships	Hardship	HHID14, UT3TYPE	6,455
B2_UT4	BK 2 Farm business (4)- crop non-land assets	Asset	HHID14, UTTYPE	80,752
B2_NT1	BK 2 Non farm business (1)-participation	Household	HHID14	15,185
B2_NT2	BK 2 Non farm business (2)-business details-grid	Business	HHID14, NT_NUM	6,740

File Name	Contents	Level of Observation	Variable(s) that Identify the Unique Observation	No. Records
B2_HR1	BK 2 Household Assets (1)-grid	Asset	HHID14, HRTYPE	197,314
B2_HR2	BK 2 household Assets (2)-transactions	Asset	HHID14, HR2TYPE	91,068
B2_HI	BK 2 household non labor income	Income source	HHID14, HITYPE	75,875
B2_ND1	BK 2 Natural Disasters (1)	Household	HHID14	15,185
B2_ND2	BK 2 Natural Disasters (2)	Disaster	HHID14, NDTYPE	3,436
B2_BH	BK 2 Borrowing	Household	HHID14	15,185
B2_TIME	BK 2 Date and time of interview	Household interview time	HHID14, TIME_OCC	
B3A_COV	BK 3A Cover (Individ Adult)	Individual	HHID14, PID14	36,400
B3A_DL1	BK 3A Education (1)	Individual	HHID14, PID14	34,470
B3A_DL2	BK 3A Education (2)	School	HHID14, PID14, DL2TYPE	33,598
B3A_DL3	BK 3A Education (3)-grid	School	HHID14, PID14, DL3TYPE	27,862
B3A_DL4	BK3A Education (4)-school details	School	HHID14, PID14, DL4TYPE	35,652
B3A_DL5	BK 3A Education (5)-expenses	Individual	HHID14, PID14	34,470
B3A_SW	BK 3A Subjective Welfare	Individual	HHID14, PID14	31,668
B3A_HR0	BK 3A Individ assets (1)-screen	Individual	HHID14, PID14	31,668
B3A_HR1	BK 3A Individ assets (2)-grid	Asset	HHID14, PID14, HR1TYPE	148,590
B3A_HR2	BK 3A Individ assets (3)-transactions-grid	Asset	HHID14, PID14 HR2TYPE	68,562
B3A_HI	BK 3A Individual non labor income	Income source	HHID14, PID14, HITYPE	158,215
B3A_KW2	BK 3A Marriage (1)-screen	Individual	HHID14, PID14	34,470
B3A_KW33	BK 3A Marriage (3)-history	Marriage	HHID14, PID14, KWN	6,108
B3A_PNA1	BK 3 A Positive/Negative Affects (1) - Activities yesterday	Individual	HHID14, PID14	34,470
B3A_PNA2	BK 3 A Positive/Negative Affects (2) - Types of Experiences	Experience	HHID14, PID14,PNATYPE	382,664
B3A_RE1	BK 3A Retirement (1)	Individual	HHID14, PID14	27,225
B3A_RE2	BK 3A Retirement			19,054
B3A_PK1	BK 3A HH decision making (1)	Individual	HHID14, PID14	24,896
B3A_PK2	BK 3A HH decision making (2)	Decision	HHID14, PID14, PK2TYPE	388,692
B3A_PK3	BK 3A HH decision making (3)	Status indicator	HHID14, PID14, PK3TYPE	71,056
B3A_BR1	BK 3A Pregnancy summary	Individual	HHID14, PID14	31,148
B3A_MG1	BK 3A Migration (1)-birthplace	Individual	HHID14, PID14	33,972
B3A_MG2	BK 3A Migration (2)-history	Migration event	HHID14, PID14, MOVENUM	19,813
B3A_SI	BK 3A Risk and Time preferences	Individual	HHID14, PID14	31,668

File Name	Contents	Level of Observation	Variable(s) that Identify the Unique Observation	No. Records
B3A_TR	BK 3A Trust	Individual	HHID14, PID14	31,668
B3A_TK1	BK 3A Work history (1)-screen	Individual	HHID14, PID14	34,470
B3A_TK2	BK 3A Work history (2)-current job	Individual	HHID14, PID14	24,494
B3A_TK3	BK 3A Work history (3)-history	Individual year	HHID14, PID14, TK28YEAR	209,792
B3A_TK4	BK 3A Work history (4)-first job	Individual	HHID14, PID14	29,427
B3A_TIME	BL 3A Date and time of interview	Individual interview time	HHID14, PID14, TIME_OCC	
B3B_COV	BK 3B Cover (Individ Adult)	Individual	HHID14, PID14	36,389
B3B_KM	BK 3B Smoking	Individual	HHID14, PID14	34,277
B3B_KK1	BK 3B Self assessed health	Individual	HHID14, PID14	34,277
B3B_KK2	BK3B Physical activities	Activity	HHID14, PID14, KKTYPE	94,407
B3B_KK3	BK3B Self-assessed health, ADL-IADL	Activity	HHID14, PID14, KK3TYPE	788,095
B3B_KK4	BK3B Self-assessed health, help	Individual	HHID14, PID14	34,265
B3B_KP	BK 3B CES-D	Question	HHID14, PID14, KPTYPE	314,530
B3B_AK1	BK 3B Health insurance	Benefit	HHID14, PID14, AKTYPE	201,427
B3B_AK3	BK3B Health Insurance	Benefit	HHID14, PID14, AK2TYPE	82,687
B3B_EH	BK 3B Early Health	Individual	HHID14, PID14	31,477
B3B_FM1	BK 3B Food and meals 1	Individual	HHID14, PID14	31,477
B3B_FM2	BK 3B Food and meals (2)	Food Type	HHID14, PID14, FMTYPE	534,174
B3B_PSN	BK 3B Personality	Personality Type	HHID14, PID14, PSNTYPE	471,780
B3B_RJ0	BK 3B Outpatient care (0)-use	Individual	HHID14, PID14	34,277
B3B_RJ1	BK 3B Outpatient care (1)-medical facilities	Health facility	HHID14, PID14, RJTYPE	55,665
B3B_RJ2	BK 3B Outpatient care (2)-purposes of visits	Individual	HHID14, PID14	31,427
B3B_RJ3	BK 3B Outpatient care (3)-health examinations	Health Examination	HHID14, PID3b, RJ24TYPE	219,982
B3B_SA	BK 3B Childhood Socio-Economic Status	Individual	HHID14, PID14	31,477
B3B_TDR	BK 3B Sleep		HHID14, PID14, tdr type is gone	314,770
B3B_MA1	BK 3B Acute morbidity	Individual	HHID14, PID14	34,277
B3B_MA2	BK 3B Morbidity-symptoms	Symptom	HHID14, PID14, MATYPE	719,187
B3B_PS	BK 3B Self-treatment	Treatment	HHID14, PID14, PSTYPE	157,120
B3B_RN1	BK 3B Hospitalization (1)-use	Individual	HHID14, PID14	34,277
B3B_RN2	BK 3B Hospitalization (2)-events	Treatment	HHID14, PID14, RN1TYPE	8,360
B3B_PM1	BK 3B Community participation (1)	Individual	HHID14, PID14	34,277

File Name	Contents	Level of Observation	Variable(s) that Identify the Unique Observation	No. Records
B3B_PM2	BK 3B Community participation (2)	Community Activities	HHID14, PID14, PM3TYPE	534,072
B3B_BA0	BK 3B Non-HH mems (1)-parents	Individual	HHID14, PID14, BA04X	68,440
B3B_BA1	BK 3B Non-HH mems (2)-transfers	Individual	HHID14, PID14	17,646
B3B_BA4	BK 3B Non-HH mems (5)-sibs (transfers)	Individual	HHID14, PID14	34,220
B3B_BA6	BK 3B Non-HH mems (7)-kids (roster)	Child	HHID14, PID14, BA63A	29,585
B3B_TF	BK 3B Transfers and Arisan	Type of transfers	HHID14, PID14, TFTYPE	136,852
B3B_CD1	BK 3B Disabilities	Individual	HHID14, PID14	34,264
B3B_CD2	BK 3B Doctor diagnoses	Chronic disease	HHID14, PID14, CD01TYPE	239,848
B3B_CD3	BK 3B Doctor diagnoses	Chronic disease	HHID14, PID14, CDTYPE	548,208
B3B_CO1	BK 3B Cognitive measures	Individual	HHID14, PID14	31,477
B3B_COB	BK 3B Cognitive measures-word recall	Individual	HHID14, PID14	31,415
B3B_EP1	BK 3B Expectations for children	Individual	HHID14, PID13	31,477
B3B_EP2	BK 3B Expectations for children	Child	HHID14, PID14, EP05, EP06	13,206
B3B_TIME	BK 3B Date and time of interview	Individual interview time	HHID14, PID13, TIME_OCC	
B4_COV	BK 4 Cover (Ever married female)	Woman	HHID14, PID14	13,062
B4_KW2	BK 4 Marriage (1) current	Woman	HHID14, PID14	12,192
B4_KW42	BK 4 Marriage (2) history	Marriage	HHID14, PID14, KWN_NUM	5,886
B4_BR	BK 4 Pregnancy summary	Woman	HHID14, PID14	12,478
B4_BA6	BK 4 Non-HH members-children	Child	HHID14, PID14, BA63A	18,619
B4_BF	BK 4 Breastfeeding (Panel resp.)	Woman	HHID14, PID14	12,192
B4_CH0	BK 4 Pregnancy history (1)	Woman	HHID14, PID14	12,478
B4_CH1	BK 4 Pregnancy history (2)	Pregnancy	HHID14, PID14, CH05	14,965
B4_BX6	BK 4 Non-HH members-children	Child	HHID14, PID14, BX63A	2,126
B4_CX1	BK 4 Contraception (1)	Method	HHID14, PID14, CXTYPE	106,150
B4_CX2	BK 4 Contraception (2)	Woman	HHID14, PID14	12,478
B4_EP1	BK 4 Expected Plans for children	Woman	HHID14, PID14	12,192
B4_EP2	BK 4 Expected plans for children	Child	HHID14, PID14, EP05	14,901
B4_TIME	BK 4 Date and time of interview	Woman interview time	HHID14, PID14, TIME_OCC	
B5_COV	BK 5 Cover (Child)	Individual	HHID14, PID14	16,199
B5_DLA1	BK 5 Child's education (1)	Individual	HHID14, PID14	15,745
B5_DLA2	BK 5 Child’s education (2)-history	School level	HHID14, PID14, DLATYPE	10,839

File Name	Contents	Level of Observation	Variable(s) that Identify the Unique Observation	No. Records
B5_DLA3	BK 5 Child's education (2)-	Individual	HHID14, PID14	15,745
B5_DLA6	BK 5 Child's education (3) –work status	Work type	HHID14, PID14, DLA2TYPE	40,444
B5_FMA1	BK 5 Child's Food and Meal (1)	Child	HHID14, PID14	15,745
B5_FMA2	BK 5 Child's Food and Meal (2)	Type of Food	HHID14, PID14,FMATYPE	258,791
B5_MAA1	BK 5 Child's acute morbidity (1)	Child	HHID14, PID14,	15,745
B5_MAA2	BK 5 Child's acute morbidity (2)	Morbidity	HHID14, PID14, MAATYPE	314,840
B5_PSA	BK 5 Self-treatment	Treatment	HHID14, PID14, PSATYPE	78,710
B5_RJA3	BK 5 Outpatient care-(4) vaccine	Individual	HHID14, PID14	15,745
B5_RNA1	BK 5 Hospitalization - (1) use	Health facility	HHID14, PID14, RNA1TYPE	3,798
B5_RNA2	BK 5 Hospitalization - (2) events	Child	HHID14, PID14, RNA2TYPE	15,745
B5_BAA	BK 5 Non HHM-parents	Parent	HHID14, PID14, BAATYPE	31,490
B5_TIME	BK 5 Date and interview time	Child interview time	HHID14, PID14, TIME_OCC	
BEF_COV	BK FE Cov	Individual	HHID14, PIDPROX	2,728
BEF_1	BK Exit Form (1)	Deceased Individual	HHID14, PID14, PIDLINK	3,006
BEF_2	BK Exit Form (2)- diagnoses	Deceased Individual	HHID14, PIDLINK, EF1TYPE	26,920
BEF_34	BK Exit Form (3)	Deceased individual's relatives	HHID14, PID14	4,988
BEF_TIME	Date and time of interview	Deceased individual interview time	HHID14, PID14, TIME_OCC	
BUS_COV	BK US Cov	Individual	HHID14, PID14	52,592
BUS_US	BK US Health Assessment	Individual	HHID14, PID14	48,148
EK_EK1	BK EK1 Math/cognitive evaluations	Individual Achievement test	HHID14, PID14	66283
EK_EK2	BK EK2 Math/cognitive evaluations	Individual Achievement test	HHID14, PID14	87725

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

File Name	Contents	Level of Observation	Variable(s) that Identify the Unique Observation	No. Records
BK1A_COV	BK1: Cover A	Community	COMMID14	320
BK1A_TIME	BK1 : A Length of Interview	Community, time of interview	COMMID14, TIME_OCC	937
BK1B_COV	BK1: Cover B	Community	COMMID14	320
BK1B_TIME	BK1 : B Length of Interview	Community, time of interview	COMMID14, TIME_OCC	912
BK1C_COV	BK1: Cover C	Community	COMMID14	320
BK1C_TIME	BK1 : C Length of Interview	Community, time of interview	COMMID14, TIME_OCC	652
BK1	BK1: Book 1	Community	COMMID14	320
BK1_A1	BK1: A Destination	Destination	COMMID14, ATYPE	2488
BK1_B	BK1: B Electricity	Elec. Source	COMMID14, BTYPE	1866
BK1_C1	BK1: C1 Water Source	Water Source	COMMID14, C0	4043
BK1_C2	BK1: C2 Toilet	Toilet Facilities	COMMID14, C16A	3421
BK1_C3	BK1: C3 Garbage	Garbage Disposals	COMMID14, C18A	2799
BK1_D1	BK1: D1 Irrigation	Irrigation	COMMID14, D1TYPE	588
BK1_D1A	BK1:D1A Rice	Rice Field Types	COMMID14, D8ATYPE	428
BK1_D2	BK1: D2 Extension Activity	Activity	COMMID14, D2TYPE	1060
BK1_D3	BK1: D3 Crop	Crop	COMMID14, D3TYPE	614
BK1_D4	BK1: D4 Factory	Factory	COMMID14, D4TYPE	611
BK1_E1	BK1: E1 Name Change	Name Change	COMMID14, E1TYPE	18
BK1_E2	BK1: E2 Major Event	Major Event	COMMID14, E10EVT	537
BK1_F1	BK1:F1 Natural Disasters	Natural Disasters	COMMID14, F1TYPE	2488
BK1_F2	BK1:F2 Damage from disasters	Damage from disasters	COMMID14, F2TYPE	2160
BK1_G	BK1: G Credit	Credit Inst.	COMMID14, GTYPE	3110
BK1_I	BK1: I History schools	School Level	COMMID14, ITYPE	NA
BK1_PAP0	BK1: Safety Net Programs	Community	COMMID14	311
BK1_PAP1	BK1: Safety Net Program	Program	COMMID14, PAP1TYPE	6220
BK1_PMKD	BK1: PMKD Activity	Activity	COMMID14, PMKDTYPE	5598

File Name	Contents	Level of Observation	Variable(s) that Identify the Unique Observation	No. Records
BK1_TR	BK1:TR Trust and conflict	Conflict type	COMMID14, TRTYPE	2177
BK_IR	BK1:IR	Respondent	COMMID14. IRTYPE	2488
BK2	BK2: Community	Community	COMMID14	320
BK2_KA3	BK2: KA3 Forest cleaning	Purpose	COMMID14, KA3TYPE	336
BK2_KA4	BK2 :KA4		COMMID14, KA4TYPE	880
BK2_LU1	BK2: LU1 Employment	Employment Type	COMMID14, S38TYPE	3110
BK2_PL	BK2: PL l Pollution	Pollution type	COMMID14, KA1TYPE	1244
BK2_ST	BK2 : ST Land Use Rights	Land use right	COMMID14, S1TYPE	3421
BK2_TIME	BK2 : Length of Interview	Community, time of interivew	COMMID14, TIME_OCC	544
INF	Informant	Informant	COMMID14, FCODE_INF	640
INF_PAP0	Informant	Informant	COMMID14, FCODE_INF	622
INF_PAP1	Informant Safety Net Programs	Informant/Poverty Program	COMMID14, FCODE_INF, PAP1TYPE	12420
INF_TR1	Informant conflict	Informant/ Conflict type	COMMID14, FCODE_INF, TRTYPE	4354
INF_TIME	Informant : Length of Interview	Informant, time of interview	COMMID14, FCODE_INF, TIME_OCC	1901
PKK	PKK: Community	PKK	COMMID14	320
PKK_I	PKK: I History Schools	School	COMMID14, ITYPE	1236
PKK_J	PKK: J History Health Facility	Facility type	COMMID14, J26TYPE	2163
PKK_J1	PKK:J1	Immunizations Facility	COMMID14, FASCODE, J25TYPE	1280
PKK_KSR	PKK: Assistance	Type	COMMID14, KSR2TYPE	1236
PKK_PM	PKK: PM Activity	Activity	COMMID14, PMTYPE	5562
PUSK	PUSK	Puskesmas	FCODE	961
PUSK_A1	PUSK: Change Experiences	Changes	FCODE, ATYPE	10560
PUSK_AKM1	PUSK: AKM1	Service type	FCODE, AKM1TYPE	7168
PUSK_AKM2	PUSK: Services Fees to JKN/KIS	Service type	FCODE, AKM2TYPE	7256
PUSK_C0	PUSK: C0 Hours of operation	Day of week	FCODE, C01	5760
PUSK_C1	PUSK: C1 Service	Service	FCODE, C1TYPE	72000
PUSK_C2	PUSK: C2 Referral Facility	Facility	FCODE, C2TYPE	2880
PUSK_C3	PUSK: C3 Laboratory Test	Test	FCODE, C3TYPE	12480

File Name	Contents	Level of Observation	Variable(s) that Identify the Unique Observation	No. Records
PUSK_C4	PUSK: C4	Visitors/Day of week	FCODE, C4TYPE	5754
PUSK_D1	PUSK: D1 Employee	Employee	FCODE, D1TYPE	3840
PUSK_D2	PUSK: D2	Employee type	FCODE, D3TYPE	17493
PUSK_D3	PUSK: D3 Employee information	Employee	FCODE, D10	not
PUSK_DM1	PUSK: Decision Maker 1	Type of decision	FCODE, DM1TYPE	9600
PUSK_DM2	PUSK: Decision Maker 2	Type of decision	FCODE, DM2TYPE	7680
PUSK_E1	PUSK: E1 Equipment	Type of Equipment	FCODE, E1TYPE	27840
PUSK_E2	PUSK: E2 Supplies	Type of instrument	FCODE, E2TYPE	15360
PUSK_G1	PUSK: Family Planning Cases	Type of method	COMMID14,FCODE, G1TYPE	
PUSK_VIG	PUSK: Vignette	Facility	FCODE	960
PUSKA_COV	PUSK A: Cover	Facility	FCODE	1271
PUSKA_TIME	PUSK A : Length of Interview	Facility, time of interview	FCODE, TIME_OCC	2437
PUSKB_COV	PUSK B: Cover	Facility	FCODE	1271
PUSKB_TIME	PUSK B : Length of Interview	Facility, time of interview	FCODE, TIME_OCC	1699
PUSKC_COV	PUSK C: Cover	Facility	FCODE	1271
PUSKC_TIME	PUSK C : Length of Interview	Facility, time of interview	FCODE, TIME_OCC	2471
PRA	PRA: Private Practice	Priv Practice	FCODE	3529
PRA_A1	PRA: Change Experiences	Changes	FCODE, ATYPE	15980
PRA_B1	PRA: B1 Opening and Closing Time	Day	FCODE, B1TYPE	11186
PRA_B2	PRA: B2 Service Availability	Service	FCODE, B2TYPE	83096
PRA_B3	PRA: B3 Referral Facility	Facility	FCODE, B3TYPE	6392
PRA_B4	PRA: B4 Laboratory Tests	Test	FCODE, B4TYPE	19164
PRA_B5	PRA: B5 Service charges	Service	FCODE, B5TYPE	5268
PRA_B6	PRA: B6 Number of patients	Day	FCODE, B6TYPE	11165
PRA_B7	PRA: B7 Service Charges	Service	FCODE, B7TYPE	2328
PRA_B8	PRA: B8 Patient Diagnosis	Diagnosis	FCODE, B8TYPE	14382
PRA_C1	PRA: C1 Health Equipment	Equipment	FCODE, C1TYPE	51136
PRA_C2	PRA: C2 Health Supplies	Instruments	FCODE, C2TYPE	30362
PRA_PH1	PRA: PH1 Pharmacy	Medicine	FCODE	1598

File Name	Contents	Level of Observation	Variable(s) that Identify the Unique Observation	No. Records
PRA_PH2	PRA: PH2 Stock of Meds	Medicine	FCODE, PHTYPE	47580
PRA_VIG	PRA: VIG Private practice Vignette	Facility	FCODE	1598
POS	Posyandu	Posyandu	FCODE	709
POS_B1	Posyandu: B1-Hlth services	Hlth service	FCODE, B1TYPE	6380
POS_B2	Posyandu: B2-FP services	FP service	FCODE, B2TYPE	3190
POS_C0	Posyandu: C0-Personnel	Worker	FCODE, CADRENUM	3688
POS_C2	Posyandu C2	Activity	FCODE, C7X	3828
POS_D1	Posyandu: D1-Hlth equipment	Equipment	FCODE, D1	11484
POS_PRP1	Posyandu: Revitalization	Years	FCODE, PRPTYPE	288
POS_SDP1	Posyandu: Posyandu Resources	Source of resources	FCODE, SDP04X	1780
POS_TIME	Posyandu : Length of Interview	Posyandu, time of interview	FCODE, TIME_OCC	
PLS	Posyandu Lansia:	Posyandu Lansia	FCODE	
PLS_B1	Posyandu Lansia: B1 Services	Service	FCODE, B1TYPE	8910
PLS_C1	Posyandu Lansia: C1 Manpower	Number of Cadre	COMMID14, FCODE, CADRENUM	2669
PLS_D1	Posyandu Lansia: D1 Equipment	Instrument	FCODE, DTYPE	4950
PLS_SDP	Posyandu Lansia: SPD Resources	Source of resources	FCODE, SPD04	2094
PLS_TIME	Posyandu Lansia : Length of Interview	Posyandu Lansia, time of interview	FCODE, TIME_OCC	732
TRA	TRA: Traditional practitioners	Traditional practice	FCODE, KRTYPE	4717
TRA_TIME	TRA: Trad. Pract. Length of Interview	Traditional practice, time of interview	FCODE, TIME_OCC	948
SCHL	SCHL: School	School	FCODE	2567
SCHL_A1	SCHL: A1 Training	Training type	FCODE, A1TYPE	7680
SCHL_A2	SCHL: A2 School feeding programs	Food type	FCODE, A2TYPE	702
SCHL_B10	SCHL: B10 School-related Decision	Decision type	FCODE_A, B10TYPE	38400
SCHL_B2	SCHL: B2 Shared Building	School type	FCODE, B2TYPE	1080
SCHL_B3	SCHL: B3 School facilities	Facility type	FCODE, B3TYPE	58880
SCHL_B5	SCHL: B5 Scholarships	Type	FCODE, B5TYPE	12800
SCHL_C1	SCHL: C1 Teacher	Teacher training	FCODE, C1TYPE	7677
SCHLC_F2	SCHL: F2 Student Test Scores	Test Scores	FCODE	60775
SCHL_D1	SCHL: D1 Class teacher descriptions	Classroom	FCODE, D1TYPE	5760

File Name	Contents	Level of Observation	Variable(s) that Identify the Unique Observation	No. Records
SCHL_D2	SCHL: D2 Class teacher descriptions	Classroom	FCODE, D2TYPE	4800
SCHL_E	SCHL: Student Expenditures	Item	FCODE, ETYPE	53760
SCHL_G1	SCHL: G1 Class size	Primary Class	FCODE, PRMGRD	5760
SCHL_G2	SCHL: G2 Class size	Secondary Class	FCODE, SCDGRADE	4800
SCHL_G3	SCHL: G3 Number of teachers at each level	Teacher education level	FCODE, G6X	15360
HPS	Prices-market	Market	FCODE	NA
HPS_H	Prices- market	Type of product	FCODE, HTYPE	NA
HPS_HOUR	Prices- market opening time	Time open	FCODE, DAY	2240
HPS_TIME	Prices-market interview time	Market, time of interview	FCODE, TIME_OCC	380
HTW	Market prices-shops	Shop	FCODE	640
HTW_H	Market prices-shops	Shop/Product type	FCODE, HTYPE	16640
HTW_HOUR	Market prices-shops opening time	Time open	FCODE, DAY	4480
HTW_TIME	Market prices-shops interview time	Shops, time of interview	FCODE, TIME_OCC	710
HIN	Market prices-informant	Informant	FCODE	320
HIN_H	Market prices-informant	Type of product	FCODE, HTYPE	13440
HIN_TIME	Market prices-informant time	Community, time of interivew	COMMID14, TIME_OCC	356
SAR	Service Availability Roster	Facility	FCODE	60522
SAR_COV	SAR: Cover	Community	COMMID14	320

Appendix C:
Module-Specific Analytic Notes

This appendix presents detailed notes about IFLS5 data from the household survey that may be of interest to analysts who will use the data. With the move to CAPI in IFLS5, more data checking was done while in the field than in the past. Still, the emphasis in post-field activities was to get public use files ready as quickly as possible. Hence data checking has been limited and is mostly up to users.

Book K: Control Book and Household Roster

Cover (BK_COV)

Some respondents listed on the cover page were not household members. 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 before 2014, but not in 2014. 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 SC (BK_SC)

1. SC01, SC02 and SC03 provide 2013 BPS codes for province, district (kabupaten) and sub-district (kecametan), respectively. These codes, which are also in HTRACK, were matched to the 2014 BPS codes after the fieldwork, using a crosswalk obtained from BPS. Careful cross-checking of both codes and names was done as part of the process to replace 2013 codes with 2014 codes in HTRACK14.

2. As explained above, the 2014 BPS codes should correspond to the February 2015 SUSENAS codes. Discrepancies may exist however. The codes are usually announced in mid-year, but in fact codes are being changed throughout the year. This means that some of the 2014 codes might have been changed before February 2015. SUSENAS public use generally does not come with location names, only codes, so it is not possible to tell easily if a mismatch has occurred. Another warning has to do with matching to PODES. In principal the 2014 codes should come close to matching those used in the 2014 PODES, which was fielded after mid-year, 2014 (in fact the 2013 BPS codes should match to the 2014 PODES). In fact we have found for the 1999 PODES and 1998 BPS codes, using a version of PODES with location names and codes, that some locations do not match both names and codes. This can happen for several reasons. First, PODES like IFLS and SUSENAS is a sample of communities, it is not a census. So there are some locations in PODES that do not appear in IFLS (or SUSENAS), and visa versa.

More disturbing, in about 10 percent of cases, one gets a match on location codes at the desa-level between desas in IFLS and PODES, but not on names. Maybe half or more of these mismatches are cases in which names are very close but spelled slightly differently; hence are essentially a match. However, about 5 percent are not a match for names, and yet the names in IFLS can be found in PODES, but with different codes than they have in BPS. Upon investigation at the BPS mapping department, it turns out that one group is responsible for codes for SUSENAS, SAKERNAS and other household surveys at BPS, while another is

responsible for PODES, and the codes used by each do not necessarily match. Note that the match at the kecametan level is better than at the desa level, and to protect privacy of respondents we only release location codes to the kecametan level, but there is still an issue here that most users of BPS data are probably unaware of.

Module AR (BK_AR0, BK_AR1)

1. For origin and IFLS split-off households, much information from the past household rosters was preloaded onto CAPI so that interviewers would know whom they were looking for and to obtain updated information on all household members from previous waves. The preloaded variables include PID97, AR01, AR02, AR00id (PIDLINK), AR07, AR08, AR08a, AR01g, AR01h and AR01i.

2. Variable AR01a indicates the household member's status in the 2014 household:

Origin and old split-off households:

0 = past member deceased in 2014

1 = past member still in 2014 household

2 = past member who returned in 2014

3 = past member who had left by 2014

5 = 2014 member not present in household in past waves (new member)

6 = Duplicate PIDLINK

11= member found during tracking, but not in household during main field survey

Finally code 6 were assigned to duplicate cases, that are described below in point 3.

3. We found in the field in 2007, that in 2000, some individuals in splitoff households who were thought to be new individuals and thus given new pidlinks, were actually panel members and should have been given their old pidlinks. After fieldwork, we reviewed these cases carefully and in some instances decided that they were indeed panel observations. There were two generic types of cases. First the index panel respondent moved together with another panel respondent (who was the person being tracked) to a new, splitoff household and the enumerators did not realize that the index panel respondent was a panel person, and so a new pidlink was given. Yet the person's name, age and sex were identical. In this case, we simply changed the pidlink in the splitoff household to the original one and changed the data, both in IFLS4 and in IFLS3.

The second type of case was the index panel respondent moving back to an original IFLS household, or a previous splitoff household where they were found in a prior wave. In this case sometimes the enumerators mistakenly thought this person was new and so the person was given a new pidlink. Again, if the name, age, sex corresponded exactly to another person in this household with a different pidlink, we checked thoroughly in the field during IFLS4 and if we found the person was really the same we later gave them the same pidlink. However, we chose not to delete the duplicate entry because doing so would change all the PIDs which would create confusion and possibly errors. Hence we kept the AR roster as is, but added a code in AR01a, 6, to indicate a duplicate observation. These PIDs should be ignored by users.

In PTRACK, however, since we do not include an entry for the duplicate pidlinks within the same household. Note however that individuals may still appear in multiple households if they have ever lived in one of them. But within one wave, AR01a will equal 1 only for the household in which that person was found in that wave.

4. In the fielded version of the survey, variables AR01g and AR01h indicated whether a respondent should be treated as a panel or new respondent in books 3 and 4, based on whether they completed books 3 or 4 in IFLS4.

5. Variable AR01i indicates whether the individual was supposed to be interviewed. In origin IFLS1 households all members were to be interviewed or proxy books gotten for them. In some instances users will find that current members in these households will not have either individual or proxy books. In split-off households, whether the split-off occurred in 1997, 1998, 2000, 2007 or 2014, all members of IFLS1 households, their spouses and biological children were supposed to be interviewed. If such persons were current members of the household, AR01i should equal 1. If they had moved, or if the member was not a panel IFLS1 member AR01i was set to 3. Occasionally a person was interviewed when they should not have been. We left the data as is for such cases. Also there are some households in which all current members have AR01i equal to 3. These are cases, usually split-off households, in which the IFLS1 members and their spouse and children have left the household.

6. For age (AR09), we leave all records as is, knowing that there always exists serious measurement error in age. As noted, in PTRACK we make our best guess for each wave for age and date of birth of each respondent.

7. Variables AR10, AR11, AR12, and AR14 provide the roster line number (PID07) of an individual's father, mother, caretaker (for children), and spouse (for married respondents), if they were members of the household. Because the preloaded rosters contained all past 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. Interviewers were instructed to enter the parent's roster PID even if the parent was no longer in the household (rather than enter code 51 for not in the household).

Book EF: Exit Interview

1. Most cases have a book K in their respective IFLS5 households, HHID14, however there are a small number that do not. These cases are ones for which all household members from 2007 died. For these, HHID14 and PID14 refer to the household number from IFLS4.

Book 1: Expenditures and Knowledge of Health Facilities

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

1. Some households reported little or no food expenditures. We believe that generally 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. In some cases there was bulk purchasing of some staples such as rice. One can detect this by noting a zero in purchases during the last week, but a large past purchase recorded in KS13b and 14.

2. Expenditure questions dealt with different reference periods: weekly, monthly, and yearly. Calculation of total expenditures requires standardizing on one reference period.

Module KSR (B1_KSR1, B1_KSR2)

We changed KSR to reflect new assistance programs put into place between IFLS4 and 5. For the food assistance programs; Raskin and Market Operations, we ask about participation in the last one year and the number of times. We also ask the quantities received, prices paid, the out of pocket spending and the market value. The difference between market value and out of pocket spending represents the subsidy.

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 12 types of facilities, 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

Module KR

KR05a asks about the rental value of houses that are owned by owner-occupiers. Note that these can be highly inaccurate if there are not rental markets in the area, as sometimes occurs in rural areas. In IFLS one sees a far higher variance in rental rates from KR05a than from KR04a, which is for renters. Means are not so different, though that is a little hard to interpret unless one is talking about the same location. It is certainly possible to impute rent for owners using hedonic price models, but then one has to make several strong assumptions that are may be wrong; such as that the prices of house characteristics are the same for renters and owners. In addition one would like ideally to allow housing characteristics to have different prices in different market areas, which may result in a small cell size problem, especially if rental markets are thin in an area or if the number of households in an area is small, as it will be for movers.

In 2007 we added two questions regarding major improvements in housing since 2000 (KR24b, c), ranging from building new structures, to replacing the kitchen, to putting in new tile or wood floors and so on. These questions were kept in IFLS5. We obtain the value of such improvements and whether they are real improvements or repairs to the house after major damage from a natural disaster. Households tend to put in major housing improvements when their incomes rise substantially. In particular the incidence of such improvements may be a good indicator of recovery from the financial crisis or from other negative shocks. During pretests, we saw such instances in a village in central Java in which many houses had put in new high quality tile flooring. We later found out that the village had gotten several new factories located there, plus an old factory began re-hiring workers after having laid off many during the financial crisis. In a neighboring village, we did not see any such improvements, in that village, there was no recent increase in the number of local factories. In COMFAS we ask about new factories and factories re-hiring in an area.

Module UT

UT asked details about the rice crop, particularly about crop production and value by season. This is to allow calculation of net production. Note that production needs to be converted into milled rice to be able to compare to rice consumption from section KS. Although there are published ratios of paddy to milled rice, these are averages. Milling ratios differ by variety of paddy, by season (particularly wet versus dry) and by characteristics of the field and cropping season (like how wet it was). For this reason we attempted to get from each rice farmer the milled production equivalent of their production. In general this will be noisier and possibly a good deal noisier than production estimates. Milling is generally done in pieces, whereas production tends to be harvested at one time. This makes it easier for farmers to remember paddy production than the milled equivalent.

On exception for the rule that production estimates are reliable may be when the farmer has sold part of the harvest (usually rice) to others to harvest. This is not uncommon in Java, where farmers will sell the rights to

harvest part of the crop to outsiders (the farmer does the work prior to harvest). In this case we still get an estimate of total production, which has two parts, the farmer's own harvest, and the harvest by others of the rest of the farmer's land. This last part is clearly an estimate. In variable UT07k we recorded the area harvested by others when the farmer has sold part of the harvest in this way so users will know.

Note that area estimates may be more inaccurate than production. When pretesting in central Java, we found some cases of reported yields of 9-10 tons per hectare, with IR-64. This is unlikely, even in central Java, the maximum yields at the International Rice Research Institute experimental fields for IR-64 are 10 tons per hectare (Dr. Keijiro Otsuka, private communication); 5-7 tons would be what is normally expected in central Java. Other provinces will have lower yields in general. We are pretty sure that the source of the high yields was measurement in area, very small plots being reported as too small.

UT07xa asks about the number of rice crops that the farmer can get in one year. This is probably the single best measure of irrigation quality in Indonesia. In much of Java farmers get 3 rice crops per year. Generally a third crop requires a farmer to have tube well irrigation, powered of course by a pump. Two crops can be had with gravity irrigation, plus rainfall. This measure still does not capture differences in drainage and keeping canals clean, but short of that may be very useful.

Module NT

NT asked different questions to assess business profits in different ways, by business (NT07, 8, 9).

Module HR

HR10 asked who owned household or "nonbusiness" assets, and HR12 asked what fractions were owned by husband and wife. HR10 in three cases identified both respondent and spouse as owners, but HR12 recorded only one of them as owner. There were also cases in HR10 identified both respondent and spouse as owners, but HR12 didn't record either one. Reports of fractions owned by husband and wife do not add up as expected in a handful of cases. 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%.

Land in HR in IFLS3 should not include farm land, since that was listed in module UT.

Book 3A: Adult Information (part 1)

Module DL

1. Several DL questions pertained to schooling, including the date of leaving school and dates various EBTANAS and UAN 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 or UAN tests, often taken directly from a score card, are believed to be more reliable.

2. The EBTANAS/UAN/UN scores in variable DL16d are not necessarily comparable across the country. 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.

3. Whenever possible, interviewers recorded EBTANAS scores from the EBTANAS score card. Otherwise, the interviewer had to rely on the respondent's recall. 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.

Module SW

There is heaping on 3 in SW01-03, the income ladder questions.

Module HR

The notes about module HR in book 2 apply to book 3A as well. Asking HR questions to other members of the household besides the respondent for Book 2 is designed to provide users with multiple estimates of assets, which are particularly noisy in most data sets. Because of a programming error, we only have unfolding brackets for rows A, B, C, G and J in this section.

Module KW

Questions KW14a–g asked both husband and wife about decisions on where and with whom to live after marrying. Sometimes 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.

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-resident 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 TK

For occupation and industry/sector we obtained open-ended answers. The open-ended answers were later coded into 2-digit ISTC codes for occupation and 1 digit sector codes. This was done by updating a "dictionary "of Bahasa Indonesia phrases created for IFLS2 and extended in IFLS3 , 4 and 5. By considerable checking and cross-checking this led to a consistent method to code occupations across the waves of IFLS. We checked to make sure that our updates did not imply changes to coded occupations in IFLS4, 3, 2 and 1. In some cases where it did, we accepted the changes and the earlier IFLS data were corrected. In other cases we did not accept the dictionary changes and we re-coded the translations. Eventually we converged to a new dictionary and set of occupation and sector codes, again that are as consistent across rounds as we could make them.

Book 3B: Adult Information (part 2)

Module BA (Parent) (B3B_BA0, B3B_BA1)

1. In the past BA data about parents' survival status and residence do not always agree with information in module AR. It was difficult to ascertain which module was 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. In IFLS4, section BA is only for biological parents. Non-biological parents are covered in section TF.

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, some 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.

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

Data are provided about the characteristics of non-resident children, both biological and step- or foster-children. Explicitly adding step- and foster-children was continued from IFLS3 and 4. Information is also asked about transfers of money, goods, or services between respondents and those children.

Linking Children in IFLS5 BA Rosters to Their IFLS1, 2, 3 and 4 Data. For panel respondents who reported children in 2007, we preloaded into CAPI the name, age, and sex of all children, biological and non-biological, alive in IFLS4. In IFLS5 interviewers used these preloaded child rosters to collect data on the same children. BA63a lists the line number of this child in IFLS3 BA. BA64a provides the age of the child in 2000 and BA64c registers whether the child lived in the household in 2007. To facilitate linking data on children in the IFLS5 BA rosters to data on those same children in IFLS1, 2, 3 and 4, we have provided the following variables:

- BAAR00 (IFLS5 household roster number)
- BA63a (line number in IFLS5 BA roster)

Any person who has ever been a household member is listed in the AR household roster. Hence if the child had been a member in 1993, 1997, 1998, 2000 or 2007, that child would be listed in the IFLS5 AR roster. From AR, one can pick off the child's PIDLINK, make sure that AR01a=1 and match backwards, or one can use HHID14 together with PID14.

Book 4: Ever-Married Woman Information

Module KW

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

Module CH (B4_CH0, B4_CH1)

Variables CH01ab, CH01ac, and CH02a summarize pregnancies since the last interview for panel respondents who were interviewed in IFLS4. Each woman who had answered book 4 in 2007 had a preload that listed her youngest child for whom IFLS had a record.

However, occasionally the CH module contains data on what appears to be the youngest child listed in the preloaded information. This also occurred in prior waves. With CAPI this was identified during the interview

and was corrected.

Book 5: Child Information

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 or a grandmother who had lived in the household in 1993, was no longer living in the household in 2014, but was deemed the most knowledgeable source of information for the child. In those cases the aunt/grandmother'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 her characteristics.

Module DLA (B5_DLA1)

1. Regarding the age at which the respondent entered elementary school, in a very few 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. [check]

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. [check]

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.

Book EK: Cognitive and Math Test

Module EK[BEK]

The first question, EK0, is a practice question and should not be counted. Each test question has an "X" variable associated with it, which indicates whether the answer is correct or not. There were two test booklets, one for children aged 7-14 and one for adults: aged 15 and above. The variable ekage indicates which version of the test was given. For panel children who were now 15-24 but were 7-14 and took EK1 in 2007, they were given both EK1 and 2 in 2014. The 7-14 year olds, who were more likely to still be in school, were given more questions: 12 cognitive and 5 math. The 15-24 year olds were only given 8 cognitive questions and 5 math questions. Those above 60 were not given math questions. This was to avoid refusals among older respondents. The question numbers are unique, so that question 6 in the 7-14 age book will be identical (except for color) to question 6 in the 15-24 year book. The first 12 questions are cognitive for both groups and the last 5 questions were simple math questions for the 7-14 age group and the last 10

questions for the 15-24 age group (see the questionnaire). As can be seen, the cognitive questions overlap for the two groups, while the math questions were more difficult for the older group.

COMFAS School Books

In IFLS5 the school book in COMFAS was divided into 4 different books (with the same questions), A, B, C and D, in order that we could interview different respondents at the same time.

Glossary

A-D	
ADL	Activities of daily Living
Apotik Hidup	The plant, usually used for traditional medicine
APPKD/PAK	Village Revenue and Expenditure/Village Budget Management
Askabi	Public assurance for acceptor of control birth
Arisan	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.
Bidan	Midwife, typically having a junior high school education and three years of midwifery training.
Bidan Desa	Midwife in village, Indonesia government's project to provide health service of maternal case in village such as; pregnancy check, delivery, contraception, etc.
bina keluarga balita	child development program.
bina keluarga remaja	youth development program
bina keluarga manula	ageing care program
BLT	*Bantuan Langsung Tunai* (Direct Cash Assistance), unconditional cash transfer program, later renamed as BLSM.
BLSM	*Bantuan Langsung Sementara Masyarakat*. (Direct Temporary Assistance for the People), unconditional cash transfer program replacing BLT.
Book	Major section of an IFLS questionnaire (e.g., book K).
BPS	Badan Pusat Statistik, Indonesia Central Bureau of Statistics.
BP3	Board of management and development of education, an school organization that has responsible on education tools supplies. Usually it consists of teachers and student's parents.
BUMN/BUMD	National committee/ Regional committee
CAFE	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.
CAPI	Computer Assisted Personal Interviewing, a system in which computers are used in administering survey questionnaires. With CAPI, interviewers use a portable computer/tablet to enter the data directly during face-to-face interview.
CES-D Scale	Center for Epidemiologic Studies Depression Scale, a short self-report scale designed to measure depressive symptoms.
CFS	IFLS Community-Facility Survey.
CHRLS	China Health and Retirement Study.
CPPS-UGM	Center for Population and Policy Studies of Gajah Mada University
CSPro	The Census and Survey Processing System (CSPro). The software used for entering, editing, tabulating, and disseminating census and survey data
COMMID	Community ID. This is a unique identifier for a community or Enumeration Area.
DBO	Operational Aids for School from Social Safety Net Program
Dana Sehat	Fund for health service that was collected from community of village to be used for the community
Dasa Wisma	A group of community per 10 houses, but practically 10-20 houses, to run Village programs
data file	File of related IFLS3 variables. For HHS data, usually linked with only one HHS questionnaire module.
DBS	Dried Blood Sample.
Desa	Rural township, village. Compare *kelurahan*.
DHS	Demographic and Health Surveys fielded in Indonesia in 1987, 1991, 1994,

	1997.
Dukun	Traditional birth attendant.

E-L	
EA	Enumeration Area.
EBTA	Regional Achievement Test, administered at the end of each school level, covered Agama, bahasa daerah, kesenian, ketrampilan, etc, exception subject of EBTANAS.
EBTANAS	Indonesian National Achievement Test, administered at the end of each school level (e.g., after grade 6 for students completing elementary school). Covered 5 subject; Bahasa Indonesia, Mathematic, PPKN, IPA, IPS
ELSA	English Longitudinal Study of Ageing.
GSOEP	German Socio-Economic Panel
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. IFLS3-HHS refers to the 2000 wave.
HRS	Health and Retirement Study fielded in the US
HTRACK	An IFLS data file indicating what data are available for households in each survey wave.
IADL	Instrumental Activities of daily Living.
IDT	Presidential Instruction on Undeveloped Village
IDUL FITRI	A muslim celebration that marks the end of fasting month, Ramadan.
IFLS	Indonesia Family Life Survey. IFLS1, IFLS2, IFLS3, IFLS4, and IFLS5 refer to the 1993, 1997, 2000, 2007 and 2014 waves, respectively. IFLS2+ refers to the 25% subsample wave in 1998.
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 *original IFLS1 release*.
interviewer check	Note in a questionnaire for the interviewer to check and record a previous response in order to follow the proper skip pattern.
JPS	Social Safety Net
JPS-BK	Social Safety Net program for Health Service
Kangkung	Leafy green vegetable, like spinach.
Kabupaten	District, political unit between a province and a *kecamatan* (no analogous unit in U.S. usage).
kartu sehat	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. The fund was from Social Safety Net program
Kecamatan	Subdistrict, political unit analogous to a U.S. county.
Kejar Paket A, Kejar Paket B	Informal School to learn reading and writing equivalent to elementary school
Kelurahan	urban township (compare *desa*).
Kepala desa	Village head
Kepala Kelurahan	Municipality Head
klinik, *klinik swasta*, *klinik umum*	Private health clinic.
Kotamadya	Urban district; urban equivalent of *kabupaten*.
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.

M-P	
Madrasah	Islamic school, generally offering both religious instruction and the same curriculum offered in public school.
Madya	Describes a *posyandu* that offers basic services and covers less than 50% of the target population. Compare *pratama*, *purnama*, and *mandiri*.
Main respondent	An IFLS1 respondent who answered an individual book (3, 4 or 5)
Mandiri	Describes a full-service *posyandu* that covers more than 50% of the target population. Compare *pratama*, *madya*, and *purnama*.
Mantri	Paramedic.
mas kawin	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).
MHAS	Mexican Health and Aging Study
MIDAS	Multiple Intelligences Developmental Assessment Scales
Mini-CFS	The miniature version of the community survey fielded in non-IFLS1 communities
Module	Topical subsection within an IFLS survey questionnaire *book*.
MxFLS	Mexican Family Life Survey
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, 2+ and 3 and contained at least one member of the IFLS1 household. Compare *split-off household*.
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 *IFLS1 re-release.*
“other” responses	Responses that did not fit specified categories in the questionnaire.
Panel respondent	Person who provided detailed individual-level data in IFLS2.
peningset	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 *mas kawin*).
perawat	Nurse.
pesantren	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; in IFLS3 called PID00.
PTRACK	An IFLS data file indicating what data/books are available for individuals in each survey wave.
PIDLINK	ID that links individual IFLS2 respondents to their data in IFLS1.
PKK	Family Welfare Group, the community women’s organization.
PODES	*Potensi Desa* (Village Potential Statistics) a data set that provides information about village/desa characteristics for all of Indonesia. The PODES is completed as part of a census of community infrastructure regularly administered by the *BPS*. Retained at village administrative offices and used as a data source for CFS book 2.
Posyandu	Integrated health service post, a community activity staffed by village volunteers.
Posyandu Lansia	Integrated health service post serving elderly. Similar to Posyandu, it is staffed by village volunteers.
praktek swasta, *praktek umum*	Private doctor in general practice.
pratama	Describes a *posyandu* that offers limited or spotty service and covers less than 50% of the target population. Compare *madya*, *purnama*, and *mandiri*.

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 *BPS*. Retained at village administrative offices and used as a data source for CFS book 2.
posyandu	Integrated health service post, a community activity staffed by village volunteers.
praktek swasta, *praktek umum*	Private doctor in general practice.
pratama	Describes a *posyandu* that offers limited or spotty service and covers less than 50% of the target population. Compare *madya*, *purnama*, and *mandiri*.
preloaded 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.
purnama	Describes a *posyandu* that provides a service level midway between a *posyandu madya* and *posyandu mandiri* and covers more than 50% of the target population. Compare *pratama*, *madya*, and *mandiri*.
puskesmas, *puskesmas pembantu*	Community health center, community health subcenter (government clinics).
RT	Sub-neighborhood.
RW	Neighborhood.

S–Z

SAR	Service Availability Roster, CFS *book*.
SD	Elementary school (*sekolah dasar*), both public and private.
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.
Sinse	Traditional practitioner.

S–Z (cont.)

SMK	Senior vocation high school (*sekolah menengah kejuruan)*.
SMP	Junior high school (*sekolah menengah pertama*), both public and private. The same meaning is conveyed by SLTP (*sekolah lanjutan tingkat pertama*).
SMU	Senior high school (*sekolah menengah umum*), both public and private. The same meaning is conveyed by SMA (*sekolah menengah atas*) and SLTA (*sekolah lanjutan tingkat atas*).
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, 2+ or 3 because it contained a target respondent. Compare *origin household*.
SPRT	Special filter paper for finger prick blood samples.
SUSENAS	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.
Tabib	Traditional practitioner.
target household	Origin household or split-off household in IFLS2 or 2+
target respondent	IFLS1 household member selected for IFLS3 either because he/she had provided detailed individual-level information in IFLS1 (i.e., was a *panel respondent*) or had been age 26 or older in IFLS1 or met other criteria, see text.
tracking status	Code in preloaded household roster indicating whether an IFLS1 household member was a *target respondent* (= 1) or not (= 3).
tukang pijat	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.
warung	Small shop or stall, generally open-air, selling foodstuffs and sometimes prepared food.

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
DL	Literacy, access to cell phone & the internet, educational level, date of school completion (or departure), EBTANAS score (for those under 30), school-related expenses.	Characteristics of schooling at each level attended (elementary, junior high school, senior high school and post-secondary).
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.
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.	Job search activities. History of jobs over the last eight years. Information on job termination and job quitting. Detailed questions on severance pay and pension benefit.
PM	Participation in an arisan, participation in community development activities.	Detail on *arisan* participation, levels and forms of participation in community development activities. Participation in elections.
KM	Whether ever smoked, what was smoked, length of time since quitting (if not a current smoker), current quantity smoked, brand and expenditures.	Details on quantity smoked and prices paid. Questions related to smoking behavior related to smoking dependence (difficulty to refrain from smoking at particular times, places, etc.)
KK	General health, physical functioning	Health expectation in the next few years and comparison with other. Physical activities in the last 7 days.
CD	Details on diagnosed chronic conditions	Same
MA	Experience of symptoms and pain in the last 4 weeks.	Additional symptoms and pain questions for individuals age 40 years old or more (e.g chest pain, cataract, glaucoma, etc).
PNA	Experience of a pain, whether the pains limit daily activities and their treatments.	Information on 11 other negative and positive feelings/affects that capture hedonic well being.
AK	Information on types of health insurance.	Same
RJ	Incidence and reasons for visits to health care providers in the past 4 weeks	Health care visits for person aged 50 years old or more. Characteristics of health care providers. Detail on services received and expenditures on care, and information on receiving health examinations.
RN	Incidence of in-patient visits to health care providers in the past 4 weeks.	Detail on services received and expenditures on care.
BR	Same as BR in Book 4	
CH	Pregnancy outcome, use of prenatal care, delivery site, survival status for up to two pregnancies in last five years.	Complete pregnancy histories. Details on prenatal services received, length of labor, birthweight, breastfeeding.
CX	Birth control methods received.	Detail on birth control device/methods
BA	Sames as BA in Book 3B.	

Module	Information in Proxy Book	Additional Information in Main Book
TF	Sames as TF in Book 3B.	

Table 2.2 Differences in Information Collected from New vs Panel Respondents in IFLS5

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. Also EBTANAS or UAN/UN test scores at each level were collected for respondents under 30. (Panel respondents who answered DLA in Book 5 in 2007 were treated as new Book 3A DL respondents).	Every level of schooling attended since June 2007 for panel respondents who had attended school since 2007. For panel respondents who answered DL in Book 3A in 2007 and did not go to school after 2007, the level of schooling ever attended. No test scores were collected.	IFLS5 can be used to create an entire history of schooling progression for respondents under 50 years
KW (marriage) Panel check: KW00a, KW02h, KW22x	All previous and current marriages.	Current or most recent marriage and any other marriage that began after 2007	For respondents who have had no marriages that ended before 2007, IFLS5 provides a complete marriage history. Data on marriages that ended before 2007 are in IFLS1, 2, 3, and 4.
MG (migration) Panel check: MG00x MG18a	Residence at birth, age 12, and all moves after age 12	All moves since residence in 2007	Use IFLS1,2, 3, and 4 for residence at birth and moves between birth and 2007.
3A, PK (household decision making) Panel check: PK 19a	Information on parents and parents-in-laws at time of most recent marriage.	Not answered for marriages before 2007.	For marriages between 1997 and 2000, use data from module PK in IFLS2 and IFLS3. For marriages between 2000 and 2007 use IFLS3 and IFLS4. A check is possible on the retrospective information by comparing IFLS4 with IFLS3, and, IFLS5 with IFLS4.

Module	New Respondents	Panel Respondents	Creating a Full History for Panel Respondents
3A, BR (pregnancy summary) book 3B Panel check: BR00xa	All live births, still births, and miscarriages (for respondents at least age 50).	Same for respondents at least 50 and not respondent of Book 4	
4, BR (pregnancy summary) book 4 Panel check: BR00x	All live births, still births, and miscarriages (new respondents and panel respondents without a child reported on preprinted child roster).	None if panel respondent had preprinted child roster with children reported.	Use IFLS1 for births up to 1993. Use IFLS2 data in CH module to compute the number of additional births from 1993 to 1997 and IFLS2+ for any births between 1997 and 1998. Use CH in IFLS3 to compute additional births between 1997 or 1998 and 2000, and, CH module in IFLS4 to compute births between 2000 and 2007.
BF (breastfeeding) Panel check: BF00	Asked in module CH (new respondent and panel respondents without a child reported on preprinted child roster).	Updates on breastfeeding for the youngest child at the last interview (IFLS4) if that child was 4 or younger in 2007 (so 11 or younger in 2014; therefore might still have been breastfeeding in 2007)	If the youngest child was still breastfeeding in 2007, use IFLS5 data in BF07 to determine the total duration of breastfeeding. For children born between 1997/1998 and 2007, breastfeeding data are in IFLS3 and IFLS4.
CH (pregnancies) Panel check CH00	All pregnancies (new respondents and panel respondents without a child listed on preprinted child roster).	Pregnancies occurring after the birth of the child who was the youngest child in 2007 (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 IFLS1,2,2+, 3, and 4.	Use the IFLS1 data in the CH module for pregnancies that began before 1993, IFLS2 for pregnancies between 1993 and 1997, and IFLS3 between 1997 and 2000. For pregnancies between 2000 and 2007, use IFLS4.

Table 3.1: Summary of weights

	IFLS1 WEIGHTS	IFLS2 WEIGHTS		IFLS3 WEIGHTS		IFLS4 WEIGHTS		IFLS5 WEIGHTS	
	Re-release Name	Longitudinal Analysis	Cross-Section Analysis	Longitudinal Analysis	Cross-Section Analysis	Longitudinal Analysis	Cross-Section Analysis	Longitudinal Analysis	Cross-Section Analysis
1	HWT93	HWT97L	HWT97X	HWT00La,b	HWT00Xa HWT00Xb	HWT07La	HWT07Xa HWT07X_	HWT14La	HWT14Xa HWT14X_
2		–	–	–	–	HWT93_97_00_07L	–	HWT_5_WAVES_L	
3	PWT93	PWT97L	PWT97X	PWT00La PWT00Lb	PWT00Xa PWT00Xb	PWT07La	PWT07Xa PWT07X_	PWT14La	PWT14Xa PWT14X_
4	PWT93IN	PWT97INL	–	–	–	PWT93_97_00_07L	–	PWT_5_WAVES_L	
5	PWT93US	PWT97USL	PWT97USX	PWT93_97_00USL	PWT00USXa PWT00USb	PWT93_97_00_07USL	PWT07USXa PWT07USX_	PWT_5_WAVES_USL	PWT14USXa PWT14USX_
6	–	–	–	–	–	–	PWT07DBSXa PWT07DBSX_	PWT14_DBSL	PWT14DBSXa

1. Household weight based on 7,224 HHs interviewed in IFLS1, all HHs interviewed in IFLS2, all HHs interviewed in IFLS3, all HHs interviewed in IFLS4, and all HHs interviewed in IFLS5
2. Household longitudinal weight for households in all full waves from IFLS1 to the latest waves
3. Person weight based on all individuals listed in a HH roster.
4. Longitudinal person weights for the IFLS1 "Main" respondents who were administered an individual book. Use these weights when using responses from “Main” respondents’ individual books (B3, B4 and B5) from IFLS1 and 2 or IFLS1, 2, 3 and 4, or IFLS1, 2, 3 , 4, and 5 in combination. There is no corresponding cross-section weight.
5. Person weights for anthropometry and health assessments in IFLS1, 2, 3, 4, and 5.
6. Person weights for DBS assayed

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 (a), or not (_). 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 (a), or not (_) and changes in the population distribution between IFLS1, 2, 3, 4, and 5. They are intended to reflect the distribution of individuals and households in the 13 IFLS provinces in Indonesia at the time of IFLS2, 3 , 4, and 5 respectively.

Table 3.2 Logit estimates of households recontact

	Contacted in 2014, conditional on contacted alive in 1993		Contacted in 1997 conditional on contacted alive in 1993		Contacted in 2000 conditional on contacted alive in 1997		Contacted in 2007 conditional on contacted alive in 2000		Contacted in 2014, conditional on contacted alive in 2007	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
ln (PCE) spline										
1st quartile	-0.229	[0.261]	-0.012	[0.160]	-0.413	[0.396]	0.081	[0.301]	0.423	[0.366]
2nd quartile	-0.55	[0.410]	-0.64	[0.411]	-1.108	[0.807]	-1.171	[0.740]	-3.229	[0.901]***
3rd quartile	-1.131	[0.300]***	-0.916	[0.343]***	0.232	[0.592]	0.722	[0.607]	1.496	[0.632]**
4th quartile	-0.588	[0.103]***	-0.294	[0.127]**	-0.479	[0.183]***	-1.207	[0.191]***	-0.972	[0.237]***
Household characteristics										
HH size	0.093	[0.030]***	0.042	[0.036]	0.01	[0.006]*	0.02	[0.005]***	0.019	[0.006]***
if 1 person HH	-0.457	[0.189]**	-1.003	[0.209]***	-1.926	[0.298]***	-1.16	[0.260]***	-0.119	[0.274]
if 2 person HH	-0.211	[0.163]	-0.51	[0.184]***	-0.632	[0.299]**	0.023	[0.266]	-0.164	[0.206]
Location										
Urban	0.574	[0.105]***	0.948	[0.126]***						
1 if all sub-hh in urban					-0.078	[0.231]	-1.091	[0.195]***	-0.947	[0.208]***
1 if sub-hh in rural and urban					1.5	[1.030]	0.521	[0.397]	0.628	[0.325]*
Province										
North Sumatra	0.607	[0.168]***	0.099	[0.198]	-0.045	[0.273]	0.027	[0.260]	0.552	[0.258]**
West Sumatra	1.268	[0.224]***	0.943	[0.265]***	1.296	[0.501]***	0.363	[0.367]	1.057	[0.369]***
South Sumatra	0.607	[0.197]***	0.279	[0.237]	0.705	[0.413]*	-0.525	[0.269]*	0.829	[0.356]**
Lampung	1.695	[0.334]***	0.786	[0.311]**	0.653	[0.480]	1.506	[0.737]**	1.647	[0.605]***
West Java	1.371	[0.155]***	1.296	[0.205]***	1.716	[0.347]***	0.956	[0.261]***	0.693	[0.210]***
Central Java	2.533	[0.263]***	2.463	[0.350]***	2.311	[0.494]***	1.074	[0.311]***	2.203	[0.408]***
Yogyakarta	1.097	[0.190]***	1.059	[0.246]***	1.16	[0.364]***	1.781	[0.444]***	0.896	[0.275]***
East Java	2.095	[0.216]***	1.183	[0.221]***	3.997	[1.025]***	1.216	[0.307]***	1.503	[0.289]***
Bali	1.554	[0.248]***	1.062	[0.283]***	2.222	[0.549]***	1.232	[0.379]***	2.952	[0.597]***
Nusa Tenggara Barat	4.629	[1.009]***	2.41	[0.476]***						
South Kalimantan	1.473	[0.256]***	0.496	[0.251]**	2.765	[1.025]***	1.259	[0.536]**	0.88	[0.374]**
South Sulawesi	1.253	[0.246]***	0.962	[0.296]***	0.603	[0.405]	0.293	[0.341]	1.552	[0.476]***
Constant	3.625	[2.558]	2.257	[1.577]	8.333	[4.774]*	2.405	[3.849]	-2.728	[4.986]
Observations	7,224		7,224		6,752		6,786		6,548	

Standard errors in brackets.
Omitted province is Jakarta.

Table 3.3 Logit estimates of individuals recontact, 1993 to 2014

	Contacted in 2014 conditional on contacted alive in 1993	
	Coeff.	s.e.
Head=1	-0.221	[0.073]***
Spouse=1	-0.184	[0.073]**
Main respondent 93	0.548	[0.040]***
Child of head=1	-0.073	[0.005]***
Age 0-9	-0.02	[0.008]***
Age 10-14	-0.021	[0.014]
Age 15-19	0.069	[0.015]***
Age 20-29	0.077	[0.009]***
Age 30-44	0.028	[0.006]***
Age 45-59	-0.028	[0.007]***
Age 60+	0.015	[0.008]*
Male	-0.146	[0.032]***
HH size=1	-1.32	[0.124]***
HH size =2	-0.777	[0.076]***
# of HHM 0-9	0.064	[0.014]***
# of HHM 10-14	0.053	[0.018]***
# of HHM 15-19	-0.007	[0.012]
# of HHM 25+	0.017	[0.016]
headeduc	-0.029	[0.004]***
Head's educ	-0.02	[0.005]***
Head's spouse educ	0.277	[0.049]***
PCE spline:		
up 3rd quartile	0.077	[0.027]***
top quartile	-0.509	[0.043]***
Interview quality		
Excellent	0.013	[0.060]
Good	0.034	[0.032]
Location		
Urban	-0.329	[0.034]***
N. Sumatra	0.222	[0.060]***
W. Sumatera	0.581	[0.073]***
S. Sumatra	0.561	[0.073]***
Lampung	0.495	[0.085]***
W.Java	0.633	[0.054]***
C. Java	1.149	[0.064]***
Yogyakarta	1.043	[0.076]***
E. Java	0.924	[0.061]***
Bali	0.868	[0.083]***
West Nusa Tenggara	1.024	[0.080]***
S. Kalimantan	0.86	[0.085]***
S. Sulawesi	0.473	[0.071]***
Constant	-0.276	[0.296]
Observations	33,081	

Standard errors in brackets. Omitted province is Jakarta

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 3.4 Logit estimates of individuals recontact - main respondents

	Contacted in 1997, conditional on contacted alive in 1993		Contacted in 2000, conditional on contacted alive in 1997		Contacted in 2007 conditional on contacted alive in 2000		Contacted in 2014 conditional on contacted alive in 2007	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
Head=1	0.295	[0.113]***	0.306	[0.129]**	1.22	[0.486]**	0.458	[0.286]
Spouse=1	0.631	[0.109]***	0.696	[0.128]***	0.853	[0.444]*	0.269	[0.268]
Child of head=1	-0.08	[0.012]***	-0.095	[0.014]***	0.178	[0.074]**	0.031	[0.045]
Age 0-9	-0.007	[0.015]	-0.001	[0.047]	-0.083	[0.169]		
Age 10-14	-0.322	[0.029]***	-0.337	[0.039]***	-0.037	[0.079]	0.245	[0.326]
Age 15-19	0.039	[0.032]	-0.161	[0.036]***	-0.088	[0.068]	0.017	[0.063]
Age 20-29	0.066	[0.015]***	0.108	[0.020]***	0.065	[0.052]	-0.009	[0.036]
Age 30-44	0.016	[0.008]**	-0.019	[0.010]*	-0.038	[0.029]	0.007	[0.020]
Age 45-59	-0.034	[0.008]***	-0.018	[0.008]**	0.003	[0.023]	-0.032	[0.017]*
Age 60+	-0.059	[0.006]***	-0.085	[0.006]***	-0.03	[0.031]	0.063	[0.028]**
Male	-0.132	[0.058]**	-0.23	[0.071]***	-0.249	[0.178]	-0.367	[0.156]**
HH size=1	-0.754	[0.141]***	-0.238	[0.173]	-0.724	[0.572]	0.159	[0.334]
HH size =2	-0.192	[0.095]**	-0.239	[0.105]**	0.148	[0.401]	0.348	[0.237]
# of HHM 0-9	0.013	[0.023]	-0.015	[0.029]	0.11	[0.086]	-0.104	[0.076]
# of HHM 10-14	0.141	[0.031]***	0.002	[0.038]	-0.055	[0.103]	-0.114	[0.094]
# of HHM 15-19	0.034	[0.022]	0.102	[0.029]***	-0.116	[0.067]*	-0.014	[0.070]
# of HHM 25+	-0.067	[0.027]**	-0.102	[0.032]***	-0.305	[0.068]***	-0.133	[0.063]**
Head's educ	-0.018	[0.007]***	-0.006	[0.008]	-0.013	[0.020]	0.001	[0.016]
Head's spouse educ	-0.008	[0.007]	-0.016	[0.009]*	-0.039	[0.022]*	0.008	[0.019]
Spouse exist	0.198	[0.082]**	0.259	[0.094]***	-0.007	[0.292]	0.215	[0.216]
PCE spline:								
up 3rd quartile	0.072	[0.039]*	0.133	[0.057]**	-0.083	[0.177]	-0.4	[0.142]***
top quartile	-0.296	[0.062]***	-0.325	[0.064]***	-0.687	[0.162]***	-0.413	[0.204]**
Interview quality								
Excellent	-0.124	[0.089]	0.086	[0.113]	0.369	[0.329]	0.116	[0.250]
Good	0.152	[0.048]***	0.002	[0.062]	0.223	[0.158]	0.096	[0.195]
Location								
Urban	-0.34	[0.052]***	-0.136	[0.061]**	-0.371	[0.190]*	-0.387	[0.130]***
N. Sumatra	0.049	[0.088]	-0.588	[0.113]***	0.322	[0.211]	1.219	[0.362]***
W. Sumatera	0.774	[0.119]***	0.015	[0.137]	2.438	[0.473]***	0.728	[0.308]**
S. Sumatra	0.248	[0.107]**	-0.074	[0.140]	1.064	[0.293]***	0.911	[0.354]**
Lampung	0.466	[0.130]***	-0.025	[0.152]	2.79	[0.731]***	1.015	[0.399]**
W.Java	0.901	[0.088]***	0.419	[0.111]***	2.338	[0.283]***	-0.024	[0.197]
C. Java	0.999	[0.096]***	0.72	[0.123]***	2.105	[0.308]***	1.755	[0.337]***
Yogyakarta	1.202	[0.122]***	1.034	[0.155]***	2.618	[0.470]***	0.8	[0.308]***
E. Java	0.748	[0.088]***	0.509	[0.118]***	2.495	[0.338]***	0.12	[0.209]
Bali & NTB	0.881	[0.098]***	0.574	[0.127]***	1.944	[0.309]***	0.84	[0.248]***
S. Kalimantan & S. Sulawesi	0.359	[0.091]***	0.122	[0.124]	2.653	[0.436]***	0.338	[0.234]
Constant	1.649	[0.441]***	2.47	[0.765]***	5.23	[2.686]*	7.163	[2.419]***
Observations	22,019		19,318		14,803		11,561	

Standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 3.5 Logit estimates of individuals recontact -roster members

	Contacted in 1997, conditional on contacted alive in 1993		Contacted in 2000, conditional on contacted alive in 1997		Contacted in 2007 conditional on contacted alive in 2000		Contacted in 2014 conditional on contacted alive in 2007	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
Head=1	0.416	[0.070]***	0.039	[0.107]	0.504	[0.143]***	0.234	[0.158]
Spouse=1	0.99	[0.073]***	0.414	[0.111]***	0.557	[0.141]***	0.026	[0.153]
Child of head=1	-0.119	[0.006]***	-0.073	[0.011]***	-0.014	[0.017]	-0.08	[0.022]***
Age 0-9	0.019	[0.011]*	0.043	[0.040]	0.014	[0.176]		
Age 10-14	-0.406	[0.017]***	-0.337	[0.035]***	-0.091	[0.091]	0.32	[0.378]
Age 15-19	-0.078	[0.014]***	-0.065	[0.024]***	-0.123	[0.070]*	-0.093	[0.078]
Age 20-29	0.129	[0.009]***	0.085	[0.015]***	-0.014	[0.028]	-0.055	[0.028]*
Age 30-44	0.029	[0.007]***	0.008	[0.010]	-0.102	[0.014]***	-0.065	[0.013]***
Age 45-59	-0.045	[0.008]***	-0.033	[0.009]***	-0.087	[0.008]***	-0.096	[0.008]***
Age 60+	-0.05	[0.006]***	-0.064	[0.006]***	-0.085	[0.006]***	-0.085	[0.006]***
Male	-0.063	[0.036]*	-0.104	[0.059]*	-0.491	[0.101]***	-0.613	[0.097]***
HH size=1	-0.819	[0.128]***	-0.603	[0.167]***	0.956	[0.260]***	0.664	[0.195]***
HH size =2	-0.283	[0.084]***	-0.358	[0.104]***	-0.014	[0.110]	0.231	[0.103]**
# of HHM 0-9	-0.042	[0.016]***	0.016	[0.027]	-0.007	[0.038]	0.005	[0.041]
# of HHM 10-14	0.031	[0.020]	0.011	[0.035]	-0.055	[0.048]	0.053	[0.052]
# of HHM 15-19	-0.01	[0.013]	0.059	[0.026]**	0.006	[0.034]	-0.006	[0.039]
# of HHM 25+	0.071	[0.018]***	-0.022	[0.028]	-0.073	[0.032]**	-0.016	[0.034]
Head's educ	-0.015	[0.005]***	-0.01	[0.007]	0.007	[0.010]	0.012	[0.009]
Head's spouse educ	-0.01	[0.006]*	-0.027	[0.008]***	-0.003	[0.011]	0.007	[0.010]
Spouse exist	0.024	[0.053]	0.207	[0.082]**	0.101	[0.110]	0.055	[0.111]
PCE spline:								
up 3rd quartile	0.065	[0.029]**	0.183	[0.054]***	0.055	[0.067]	0.083	[0.069]
top quartile	-0.257	[0.049]***	-0.267	[0.060]***	0.048	[0.123]	0.117	[0.154]
Interview quality								
Excellent	-0.086	[0.066]	0.335	[0.112]***	-0.075	[0.157]	0.112	[0.129]
Good	0.037	[0.035]	0.177	[0.057]***	0.004	[0.070]	0.039	[0.104]
Location								
Urban	-0.101	[0.038]***	0.084	[0.060]	-0.096	[0.070]	-0.359	[0.067]***
N. Sumatra	-0.376	[0.069]***	-0.341	[0.099]***	0.083	[0.167]	-0.258	[0.163]
W. Sumatera	0.274	[0.086]***	0.474	[0.131]***	-0.046	[0.171]	-0.315	[0.169]*
S. Sumatra	-0.159	[0.082]*	0.587	[0.140]***	0.172	[0.189]	-0.119	[0.187]
Lampung	-0.08	[0.095]	0.05	[0.139]	0.237	[0.206]	0.375	[0.215]*
W.Java	0.422	[0.066]***	0.671	[0.099]***	0.075	[0.138]	-0.072	[0.141]
C. Java	0.297	[0.071]***	0.598	[0.108]***	0.173	[0.147]	0.047	[0.150]
Yogyakarta	0.416	[0.085]***	1.223	[0.153]***	0.61	[0.170]***	0.447	[0.170]***
E. Java	0.287	[0.070]***	0.722	[0.110]***	0.233	[0.145]	0.087	[0.146]
Bali & NTB	0.217	[0.073]***	0.671	[0.116]***	0.433	[0.159]***	0.226	[0.152]
S. Kalimantan * S. Sulawesi	0.062	[0.071]	0.359	[0.113]***	0.032	[0.154]	-0.114	[0.152]
Constant	2.135	[0.326]***	1.102	[0.696]	4.62	[1.767]***	2.611	[1.982]
Observations	33,081		27,236		22,835		18,865	

Standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 3.6 Logit estimates of individuals recontact - health measurements

	Measured in 1997, conditional on measured in 1993		Measured in 2000, conditional on measured in 1997		Measured in 2007 conditional on measured in 2000		Measured in 2014 conditional on measured in 2007	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
Head=1	0.298	[0.084]***	0.076	[0.139]	0.317	[0.106]***	0.521	[0.099]***
Spouse=1	0.91	[0.084]***	0.417	[0.138]***	0.357	[0.105]***	0.366	[0.095]***
Child of head=1	-0.094	[0.008]***	-0.069	[0.014]***	-0.027	[0.010]***	-0.036	[0.013]***
Age 0-9	0.016	[0.010]*	-0.066	[0.037]*	-0.2	[0.050]***		
Age 10-14	-0.31	[0.022]***	-0.232	[0.032]***	-0.114	[0.023]***	0.329	[0.119]***
Age 15-19	-0.014	[0.027]	-0.127	[0.036]***	0.03	[0.023]	0.019	[0.022]
Age 20-29	0.064	[0.012]***	0.125	[0.021]***	0.081	[0.016]***	0.023	[0.015]
Age 30-44	0.016	[0.006]***	-0.019	[0.011]*	-0.022	[0.010]**	0.006	[0.009]
Age 45-59	-0.025	[0.006]***	-0.017	[0.009]*	-0.062	[0.007]***	-0.054	[0.007]***
Age 60+	-0.049	[0.006]***	-0.056	[0.007]***	-0.085	[0.006]***	-0.093	[0.006]***
Male	-0.199	[0.043]***	-0.37	[0.071]***	-0.373	[0.052]***	-0.435	[0.056]***
HH size=1	-0.49	[0.128]***	-0.295	[0.182]	-0.504	[0.133]***	-0.041	[0.120]
HH size =2	-0.167	[0.081]**	-0.091	[0.113]	-0.176	[0.085]**	0.059	[0.079]
# of HHM 0-9	-0.039	[0.017]**	0.015	[0.031]	0.01	[0.026]	-0.003	[0.030]
# of HHM 10-14	0.032	[0.022]	0.023	[0.039]	-0.039	[0.032]	-0.022	[0.038]
# of HHM 15-19	0.019	[0.017]	0.087	[0.030]***	0.048	[0.023]**	-0.025	[0.028]
# of HHM 25+	-0.071	[0.020]***	-0.037	[0.034]	-0.031	[0.025]	-0.048	[0.026]*
Head's educ	-0.012	[0.005]**	-0.013	[0.008]	-0.003	[0.006]	0.002	[0.006]
Head's spouse educ	-0.006	[0.006]	0.001	[0.009]	-0.018	[0.007]**	-0.012	[0.007]*
Spouse exist	0.146	[0.063]**	0.163	[0.096]*	0.333	[0.073]***	0.205	[0.077]***
PCE spline:								
up 3rd quartile	0.155	[0.030]***	0.108	[0.059]*	0.145	[0.047]***	-0.087	[0.051]*
top quartile	-0.446	[0.052]***	-0.252	[0.071]***	-0.608	[0.069]***	-0.475	[0.087]***
Interview quality								
Excellent	-0.095	[0.072]	0.164	[0.117]	0.098	[0.105]	0.261	[0.094]***
Good	0.107	[0.037]***	0.109	[0.063]*	0.108	[0.048]**	0.025	[0.075]
Location								
Urban	-0.114	[0.040]***	-0.013	[0.063]	-0.156	[0.048]***	-0.317	[0.048]***
N. Sumatra	-0.16	[0.071]**	-0.335	[0.126]***	-0.023	[0.098]	0.354	[0.117]***
W. Sumatera	0.515	[0.086]***	-0.081	[0.141]	0.522	[0.110]***	0.35	[0.116]***
S. Sumatra	0.248	[0.085]***	0.067	[0.151]	0.646	[0.117]***	0.389	[0.125]***
Lampung	0.636	[0.105]***	0.014	[0.161]	0.693	[0.132]***	0.474	[0.136]***
W.Java	0.71	[0.070]***	0.421	[0.121]***	0.698	[0.087]***	0.085	[0.090]
C. Java	1.315	[0.081]***	0.564	[0.127]***	0.882	[0.093]***	0.77	[0.101]***
Yogyakarta	1.322	[0.101]***	1.152	[0.172]***	1.16	[0.114]***	0.754	[0.117]***
E. Java	0.868	[0.073]***	0.476	[0.126]***	0.883	[0.091]***	0.322	[0.094]***
Bali & NTB	1.012	[0.079]***	0.247	[0.127]*	0.841	[0.096]***	0.704	[0.100]***
S. Kalimantan * S. Sulawesi	0.145	[0.072]**	0.032	[0.132]	0.441	[0.098]***	0.334	[0.104]***
Constant	0.128	[0.334]	2.61	[0.754]***	1.587	[0.734]**	0.676	[0.875]
Observations	24,479		19,440		17838		14,538	

Standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 3.6 Logit estimates of DBS collection and assay, 2007

	Contacted in 2007 conditional on contacted in 1993		DBS collected conditional on DBS target in 2007		DBS assayed in conditional on DBS collected in 2007	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
Head=1	-0.05	[0.079]	0.626	[0.070]***	-0.335	[0.082]***
Spouse=1	0.207	[0.079]***	0.828	[0.073]***	-0.026	[0.079]
Main respondent 93	-0.086	[0.005]***	-0.069	[0.007]***	0.066	[0.009]***
Child of head=1	0.413	[0.043]***				
Age 0-9	-0.112	[0.009]***	0.3	[0.011]***	0.087	[0.013]***
Age 10-14	-0.052	[0.015]***	-0.114	[0.024]***	0.012	[0.020]
Age 15-19	0.046	[0.015]***	-0.054	[0.023]**	0.17	[0.020]***
Age 20-29	0.079	[0.009]***	-0.022	[0.011]**	-0.094	[0.010]***
Age 30-44	0.034	[0.007]***	-0.029	[0.006]***	-0.022	[0.006]***
Age 45-59	-0.006	[0.008]	-0.028	[0.006]***	0.43	[0.016]***
Age 60+	0	[0.008]	0.015	[0.006]***	-0.079	[0.021]***
Male	-0.067	[0.034]**	-0.101	[0.040]**	0.004	[0.043]
HH size=1	-1.453	[0.130]***	0.181	[0.143]	0.167	[0.136]
HH size =2	-0.653	[0.085]***	0.082	[0.082]	0.137	[0.084]
# of HHM 0-9	0.019	[0.015]	0.046	[0.021]**	-0.181	[0.023]***
# of HHM 10-14	0.081	[0.019]***	0.089	[0.029]***	0.004	[0.029]
# of HHM 15-19	-0.013	[0.012]	-0.147	[0.020]***	-0.04	[0.023]*
# of HHM 25+	0.037	[0.018]**	-0.135	[0.017]***	0.684	[0.025]***
headeduc	-0.02	[0.005]***	-0.046	[0.005]***	0.008	[0.005]
Head's educ	-0.022	[0.005]***	-0.044	[0.005]***	-0.032	[0.006]***
Head's spouse educ	0.228	[0.053]***	0.489	[0.063]***	-0.362	[0.069]***
PCE spline:						
up 3rd quartile	0.12	[0.028]***	0.196	[0.040]***	-0.14	[0.042]***
top quartile						
Interview quality	-0.593	[0.045]***	-0.053	[0.081]	-0.016	[0.084]
Excellent	0.081	[0.066]	1.775	[0.065]***	0.242	[0.079]***
Good	0.117	[0.035]***	1.511	[0.047]***	0.121	[0.067]*
Location						
Urban	-0.153	[0.037]***	-0.211	[0.039]***	0.04	[0.039]
N. Sumatra	-0.16	[0.063]**	-0.209	[0.083]**	-0.215	[0.107]**
W. Sumatera	0.603	[0.080]***	0.782	[0.092]***	0.194	[0.102]*
S. Sumatra	0.434	[0.078]***	0.324	[0.092]***	-0.117	[0.108]
Lampung	0.284	[0.091]***	0.668	[0.109]***	0.218	[0.111]**
W.Java	0.771	[0.061]***	0.688	[0.073]***	0.026	[0.086]
C. Java	0.856	[0.068]***	0.586	[0.081]***	0.089	[0.093]
Yogyakarta	1.007	[0.085]***	0.75	[0.095]***	0.534	[0.109]***
E. Java	0.889	[0.067]***	0.468	[0.075]***	0.111	[0.091]
Bali	0.592	[0.088]***	0.703	[0.101]***	0.409	[0.108]***
West Nusa Tenggara	0.548	[0.081]***	0.362	[0.094]***	-0.192	[0.105]*
S. Kalimantan	1.011	[0.099]***	0.304	[0.099]***	-0.008	[0.110]
S. Sulawesi	0.248	[0.075]***	0.124	[0.089]	0.357	[0.107]***
Constant	0.507	[0.315]	-3.977	[0.524]***	-0.664	[0.564]
Observations	33,081		24,667		19,110	

Standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%