# STUDY OF THE TSUNAMI AFTERMATH AND RECOVERY (STAR):
## Study Design and Results

**Elizabeth Frankenberg**

**Peter Katz**

**Iip U'mar Rifai**

**Bondan Sikoki**

**Cecep Sumantri**

**Ni Wayan Suriastini**

**Duncan Thomas**

**This Version:** 22 May 2025

**ACKNOWLEDGEMENTS**

# Table of Contents

## I. Introduction

The Sumatran-Andaman earthquake of December 26, 2004 and the subsequent tsunami constitute one of the most powerful and deadly disasters in world history. The damage in the Indonesian province of Aceh was extreme. Some 160,000 individuals perished and an estimated 4.5 billion dollars of property was destroyed. By 2007, efforts to repair the vast damage and rebuild destroyed infrastructure constituted one of the largest projects ever undertaken in a developing country setting.

To provide evidence on the consequences of the disaster and the evolution of the recovery effort, our team fielded multiple surveys in Aceh and North Sumatra as part of the Study of the Tsunami Aftermath and Recovery (STAR). Fieldwork for the first of these surveys, STAR1, began in May, 2005, only five months after the tsunami occurred.

STAR is designed to provide data for studying many behaviors and outcomes associated with the tsunami and the subsequent recovery. The survey collects information at the individual and household levels, including multiple indicators of economic well-being (consumption, income, and assets); education, migration, and labor market outcomes; marriage, fertility, and contraceptive use; health status and use of health care; relationships among coresident and non-coresident family members; transfers among family members and inter-generational mobility; access to public services, and participation in community activities.

With these data, the analyst can address questions regarding the impact of the tsunami and post-disaster recovery on the lives of the respondents and document the effects of social, economic, and environmental change on the population.

STAR surveyed respondents living in Aceh and North Sumatra, across 13 *kabupaten*[1] with coastlines potentially vulnerable to inundation from the 2004 tsunami. The target population was located in Aceh and in the neighboring province of North Sumatra. STAR is designed to be representative at the *kabupaten* level. The original frame for the STAR survey was the 2004 SUSENAS, collected by the Indonesian Central Bureau of Statistics. *Kabupaten* were chosen to cover both the west and east coasts of Aceh as well as vulnerable areas of North Sumatra. Clusters for the survey include areas with light, medium, and heavy damage. No geographic identifiers are provided in the STAR data apart from province.

The first STAR survey (STAR1) targeted 7,157 households containing 28,376 individuals for interview. We fielded an additional four surveys over the subsequent four years and a fifth in 2014-15. Details regarding survey results for those additional waves will be reported in subsequent addendums to this document as those waves are publicly released.

---

[1] Indonesia is a presidential democracy with four levels of government.  The provincial level is just below the national level, next are *kabupaten* which are equivalent to districts, followed by *kecematan,* or subdistricts. The smallest unit is the *desa* or *kelurahan* which are villages and townships/wards, respectively.

Across all survey rounds we succeeded in obtaining at least one interview for over 95% of target respondents. Because we follow individuals who were members of an original household in the first wave (panel respondents) but who have moved out of their original household, we have added "split-off" households over time. In addition, we interview individuals who are new members of the households and we track the biological children of panel respondents as they move out of their parents' households. Outcomes for STAR1 target households are depicted in Figure 1.

**FIGURE 1**

```
┌─────────────────────────────┐                ┌─────────────────────────────┐
│  7,157 Original Households   │ ─────────────▶ │  6,703 Household Interviews  │
│                             │                │  conducted                   │
└─────────────────────────────┘                └─────────────────────────────┘
        │        │        │
        ▼        ▼        ▼
┌──────────┐ ┌──────────┐ ┌──────────┐
│ 253 HH   │ │ 154 HH   │ │ 47 HH    │
│ with no  │ │ moved and│ │ refused  │
│ contact  │ │ could not│ │          │
│          │ │ be found │ │          │
└──────────┘ └──────────┘ └──────────┘
```
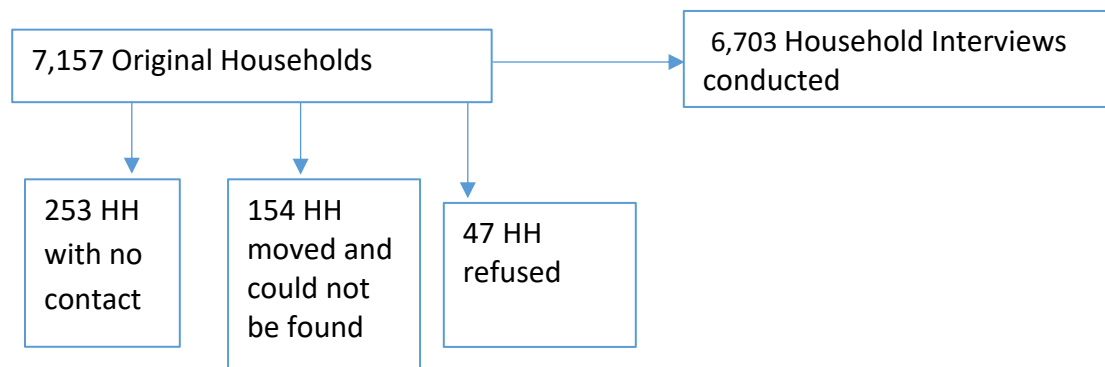
It is important to note that although STAR shares a number of common features with the Indonesian Family Life Survey (IFLS), particularly with respect to the design of the questionnaire, the two projects are different on some key dimensions. First, with respect to the populations the surveys represent, IFLS was designed to represent the Indonesian population in 1993, whereas STAR is representative of *kabupatens* in 2004. Second, the IFLS is a multi-purpose survey designed to provide data on a wide array of topics, whereas STAR is designed to measure the impact of the 2004 earthquake and tsunami and on health, economic, and social well-being.

## 2. Survey Instruments

STAR is a comprehensive multipurpose survey that collects data at both the household and individual levels. In STAR, one or two household members provide information at the household level. In addition, we attempted to conduct an individual interview with every household member aged 11 and older, and with a parent or caretaker on behalf of children 10 and younger. Because obtaining interviews with all household members is difficult, STAR includes a proxy book to collect more limited information from a proxy respondent about individuals who were not available for an in-person interview. Table 2.1 outlines the questionnaire structure and content, which are described in more detail below and in Appendix A.

The questionnaire was divided into *books* (usually addressed to different respondents) and subdivided into topical *modules*. Three books collected information at the household level, generally from the

household head or spouse[2]: book K, book 1, and book 2. Additional books collected individual-level data from adult respondents (book 3 and, after STAR1, book 3s, and book 4), and children younger than 15 (book 5). In STAR2, a health measurements module (Book US) administered to all respondents was added. Some modules appear in both a household book and an individual book (for example HR, on assets), because we wanted to collect the data both for the household as a whole and from individuals.

**Book K: Control Book and Household Roster.** Book K recorded whether a household was found and interviewed and the location of the household. If at least one member of the target household was found, information on the activities of all current members of the target household was collected.

**Book 1: Household Expenditures and Housing Attributes.** This book was typically addressed to a female respondent, either the household head, the spouse of the household head, or another person knowledgeable about household affairs. The first module recorded information about household expenditures. The second module collected information about quantities and purchase prices of several staples. The third module collected information about basic housing characteristics. The fourth module recorded information about financial and non-financial transfers into and out of the household.

**Book 2: Household Economy.** This book was usually answered by the household head or the head's spouse. Modules asked about household businesses (farm and non-farm), business assets, household assets, borrowing, and income. Combined with individual-level data on labor and non-labor income collected in book 3, this information can be used to provide a picture of current household income from market-wage income, self-employment income, family businesses, informal-sector activities, and non-labor income.

**Book 3: Individual Adult Information.** This book asked all household members 15 years and older about their education and employment status, asset ownership, marriage and fertility, and migration histories. In addition, the book included questions on morbidities, health status, health diagnoses, psychological health, usual practices (such as smoking and fruit/vegetable consumption), outpatient utilization, attitudes toward time and risk, social networks, and community participation. Finally, this book included modules on individual economic shocks and exposure to the earthquake and tsunami.

**Book 3P: Individual Adult Information by Proxy**. The proxy book was designed to facilitate collecting data by proxy about individuals who could not be interviewed directly. The proxy book contains shortened versions of most of the sections included in Book 3, excluding sections that cannot be accurately answered by a proxy respondent. In each STAR round we have combined the data from Book 3 and Book 3P to make it easier to use the data together (b3_cov contains the variable "proxy" which indicates whether a respondent answered directly, or a proxy provided answers to the questions).

**Book 3S** (after STAR1): In survey rounds after STAR1 an additional book addressed to individuals 15 and older asked more detailed questions regarding interactions with non-coresident kin (parents, siblings, and children), as well as standard batteries of questions related to post-traumatic stress and depression.

---

[2] In STAR, one member of the household was designated the household head by the person who provided information on the composition of the household. Where a married couple headed the household, the husband was generally designated the head and the wife was generally designated the spouse of the head. The head of the household is permitted to change from one STAR survey to the next, according to respondent reports.

**Book 3B** (after STAR2): In survey rounds after STAR2 a book for first-time adult respondents was created so that we could collect retrospective information for these respondents, without repeating life histories for panel respondents.

**Book 4** (after STAR1): In survey rounds after STAR1 Book 4 was administered to ever-married women 15-49 to collect information on fertility preferences and  full pregnancy histories and, for married women, current use of contraception.

**Book 5: Child Information.** This book collected information about children younger than 15. For children younger than 11, the child's mother, guardian, or caretaker answered the questions. Children between the ages of 11 and 14 were allowed to respond for themselves if they felt comfortable doing so. The five modules focused on the child's education status, work and time use, morbidities, self-treatment, and inpatient and outpatient visits. Each paralleled a module in the adult questionnaire (book 3), with some age-appropriate modifications. For example, the list of acute health conditions specified conditions relevant to younger children.

**Book US: Health Measurements** (after STAR1). This book collected physical health assessments.  These included each household member's height, weight, blood pressure, and hemoglobin status, among other measures.


## 3. STAR File Structure and Naming Conventions

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

STAR 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 KP (psychosocial well-being) collected information at the individual level (on use of counselling services) and at the symptom level, so two data files are associated with that module.

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

| Bxx_xxx | **X** |
|---|---|
| Book_Module | File sequence |


### Identifiers and Level of Observation

Wherever possible the data have been organized so that the level of observation within a file is either the household or the individual. Here we describe the conventions for STAR1, but these hold for later waves as well. If the level of observation is the household, variable IDHH01 uniquely identifies an

observation. If the level of observation is the individual, both IDHH01 and IDP01 are required to uniquely identify a person.[3]

In STAR1, IDHH01 is a nine digit character variable whose digits carry the meanings described below.

| x | x | x | x | x | x | x | x | x |
|---|---|---|---|---|---|---|---|---|
| CLUSTER | | | HOUSEHOLD IDENTIFIER | | | | | |
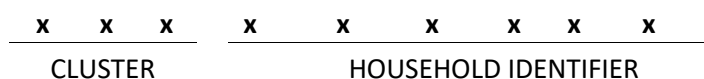
The first 3 digits correspond to a survey-defined cluster.[4] The next five digits identify a specific household. The first three digits of the household identifier are consistent across all origin households; that is, all households arising from a STAR1 target household will begin with the same three digits. The fourth through sixth digits are the household's sequence number within a cluster. The seventh and eighth digits denote the survey round at which a household is first identified – for example, "01" for a household targeted in STAR1, or "02" for a household identified as a STAR2 split-off. The final digit designates the order in which a new split-off household is found in a particular wave. For example, a household that was intact in STAR1 may have generated two new households by STAR2 if two children moved out and each form his or her own household. These new households would have a $9^{th}$ digit of 1 or 2.

The person identifier IDP01 is simply the line number of the person in the AR roster, which is two digits. In STAR, household identifiers are stored as alphanumerics, while person identifiers are stored as numerics. The alphanumeric IDPLINK identifies individuals across households. IDPLINK consists of IDHH and IDP, based on the household and roster number in which the individual was first interviewed and is constant across waves and across whatever households an individual is interviewed in. The conventions described above for STAR1 also characterize later waves of STAR.

In some data files the level of observation is something other than the household or individual. Usually this occurs 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 business asset data from module HU, file B2_HU, each observation corresponds to a particular type of asset, and there are multiple observations per household. In this data file in STAR1, for example, the combination of IDHH01 and HUTYPE uniquely identifies an observation.

The variable that defines the items or events is usually named XXTYPE, where XX 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 B2_HU above contains data about 9 asset types for each household. In other cases, the number of records per household or individual varies. For example, the level of observation in file B3_MG2 is an individual migration by a respondent. Not all respondents migrated the same number of times, so some individuals

---

[3] Within STAR1 files, use IDHH01 and IDP01 to identify individuals. Use IDPLINK to identify individuals across survey rounds. Note that in the STAR AR roster in waves after STAR1, the variable IDPLINK does not uniquely identify individuals because individuals can be listed in more than one household roster (although they are a current member of only one household).

[4] Clusters are groups of households from ecologically similar areas with respect to degree of tsunami damage, distance to the coast, elevation, and rural-urban status.

appear only once, others appear twice, and some appear more than twice. Those who did not migrate do not appear at all. This file is not rectangular because the number of observations per person is not constant. To uniquely identify an observation in this file in STAR1, for example, the analyst should use IDHH01, IDP01, and MOVENUM.

## Combining Data Across Files

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

## Concatenating Data

The analyst may wish to pool observations by concatenating two data files. For example, B3_RJ and B5_RJA both contain data on visits to outpatient providers. The data in B3_RJ pertain to adults, and the data in B5_RJA pertain to children. The variables for adults begin with RJ, while the variables for children begin with RJA, but otherwise the information is the same. In some contexts it may be useful to combine the data for the two age groups, rather than keeping it in two separate files. The data can be combined into one file using the APPEND statement in STATA. The resulting file will contain both the observations for children and the observations for adults. Because the variable names are different, the variables in one file should be renamed so that they match the names in the other file before appending.

## One-to-One Merges at the Individual or Household Level

In many cases the analyst will want to link data from one file with data about the same respondents from another file. If both files contain data at the same level of observation, the linkage will be a "one-to-one" merge. Some examples of variables to use when merging are below.

**Merging Two Files at the Individual Level of Observation.** Suppose the goal is to create a file that contains information on an individual's literacy and his or her primary activity in the past week. The file B3_DL contains information on whether respondents can read or write. The file B3_TK contains information on the respondents' primary activity in the past week. Both files contain one observation per individual. To link the desired information, sort each of the two files by person identifiers (for example, IDHH01 and IDP01 in STAR1) and then merge on those person identifiers ( IDHH01 and IDP01 in the STAR1 example).

**Merging Two Files at the Household Level of Observation.** For two household-level files, such as B1_KR and B2_UT0, which contain data on housing characteristics and borrowing, sort each file by the household identifier ( IDHH01 in STAR1, for example) and merge on that household identifier.

## One-to-Many Merges

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

This is easiest to explain using an example with STAR1 data. Suppose the analyst wishes to merge information on literacy with information on asset ownership. The identifying variables in B3_DL are IDHH01 and IDP01. The identifying variables in B3_HR1 are IDHH01, IDP01, and HR1TYPE. In B3_HR1, an individual has 7 records, one for each asset type about which we inquire. The data can be merged in two ways.

First, because the identifying variables in B3_DL are a subset of the identifying variables in B3_HR1, you could simply merge on IDHH01 and IDP01. This merge yields 7 records for each individual. Each record contains information about the individual's literacy and information about a particular asset type.

The other option is to restructure B3_HR1 so that it is organized at the level of the individual rather than at the level of the asset, in which case the identifying variables will be IDHH01 and IDP01 in the reorganized HR data. This would involve creating a file that contained variables HR01–HR07 for asset type A (i.e., HR01A–HR07A), as well as variables HR01–HR12 for the other asset types (HR01B–HR12B, HR01C–HR12C, etc.). This file would have many more variables than B3_HR1 but many fewer observations. If the data from the B3_HR1 file are restructured so that they are at the level of the individual, merging the restructured file by IDHH01 and IDP01 with the B3_DL1 data yields one record per person that contains literacy information and all information on the different types of assets.

Restructuring data files so that they are organized at a different level of observation can be done relatively easily in STATA with the reshape commands. Some data files cannot be merged without restructuring one of the data files. For example, in STAR1, the identifying variables in B2_UT1 are IDHH01 and UTTYPE. The identifying variables in B2_NT1 are IDHH01 and NT1TYPE. Neither file's identifying variables is a subset of the other's identifying variables. To merge data from these two files, you must first restructure one or both of them so that IDHH01 is the identifying variable. Generally, it is not wise to merge two files that both contain data from grids and have a TYPE identifying variable without restructuring the data.

## Questions Numbers and Variable Names

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

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 TK15 asked about respondent earnings in a calendar month. Variable TK15x indicates whether the respondent was able to answer the question. Variable TK15 provides the amount of earnings the respondent reported. 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 TK15x, 1 if a valid amount is provided, 3 if the amount is nothing, 7 if the respondent refused, 8 if the respondent doesn't know the amount, and 9 if the response is missing). 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 STAR questions required either a number or a closed-ended categorical response; a few questions allowed an open-ended response.

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

Questions requiring categorical responses usually allowed only one answer (for example, "Was the school you attended government, private, or pesantren?"). 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, "Why did you leave school?"). 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 DL07 asked why the respondent left/missed school and allowed up to 15 reasons in response. The codebook shows how many respondents cited illness and how many respondents cited marriage but not how many respondents cited both illness and marriage.

Additional response categories were sometimes added in the process of cleaning "other" variables. Typically these categories were added below the existing "other" category. For example, question TK29 asked about why a respondent was unemployed. The questionnaire was fielded providing one substantive choice ("tsunami") and a second "other". When the "other" responses were reviewed, seven other additional and commonly identified categories were added.

### Missing Values

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

For many variables, we can distinguish between *system missing data* (data properly absent because of skip patterns in the questionnaire) and data missing because of interviewer error. The data entry software generated some missing values automatically as a result of skip patterns. For example, question DL00 in book 3 asked whether the respondent had ever attended school, and if not, interviewers skipped to question DL12. If the interviewer recorded 3 (No) for question DL00, during data entry the software automatically skipped to DL12 and filled the book 3 DL02-DL11 variables with a period or blank. If data were missing because the interviewer neglected to ask the question or fill in the response, the data-entry editor was forced to enter 9 or Z in the data fields in order to get to the questions that the interviewer did ask.

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

### Special Codes and X Variables

Many STAR 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; labeled "Top Code" when this occurs in x variables
>
> 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, for most variables 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 KH06 (how much the respondent paid for an amount of food) is KH06X. The value of KH06X is 1 if the respondent provided a valid numeric answer and 8 if the respondent did not know what price he or she paid.

An indicator variable sometimes reveals more than whether special codes were used. For example, for KR12 (duration lived in house of interview), KR12X indicates both the units in which duration was recorded (weeks, months, or years) and the existence of valid numeric data.

### 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; HR (household assets) data, where the level of observation is a type of asset; and KP (psychological health) data, where the level of observation is a symptom. The name of the variable that identifies the particular observation level typically contains the module plus "TYPE," e.g., HRTYPE. 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.

### Weights

HTRACK contains weight variables (HWT and HWTP) for the given round. HWT is raked to Statistics Indonesia household data from the relevant survey year (e.g., 2005 for STAR1, 2006 for STAR2) for the STAR districts, based on district, urban-rural status, household size, and age of the household head. HWTP scales households so that they sum to the number of households with the STAR districts in that relevant survey year (e.g. 2005 for STAR1, 2006 for STAR2).

PTRACK contains weight variables (PWT and PWTP) for the given round. PWT is raked to Statistics Indonesia population data from the relevant survey year (e.g., 2005 for STAR1, 2006 for STAR2) for the STAR districts, based on age, sex, district, and urban-rural status. PWTP scales individuals so that the sample as a whole equals the estimated number of individuals within the STAR districts in the relevant survey year (e.g., 2005 for STAR1, 2006 for STAR2).

### Privacy-Protected Information

In compliance with regulations governing the appropriate treatment of human subjects, information that could be used to identify respondents in the STAR survey has been suppressed. The types of information include location, lengthy text responses, and very rare response codes. Interview dates, which are provided in HTRACK, have been randomized across a period within about two months of the actual interview date.

## 4. Special Features of the STAR Data

This section discusses the distinctive features of STAR1 data as they affect analysis files.

### Symmetric Information

In some STAR modules, respondents within a household provided symmetric information. These data may permit comparisons that help validate or fill in incomplete sections in the data. In STAR1, for example,in module KF, individuals provided information about the dates of their marriages. Within a household, if two individuals are married to each other, the KF data from the two individuals could be compared for consistency.(In STAR2 onward, one would use the KW section to compare reports on the same marriage.) Or, if one individual's data is missing, data from the spouse could be used to fill the gap. Similarly, in module MG, individuals described their migration experiences. If couples or parent-child pairs had moved together and each individual answered MG, their responses could be compared to check consistency, or the MG data of one could supply information missing from the other.

### Duplicate Information

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

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

**Birthdate.** Information on birthdate was collected in individual books and recorded in the AR roster. We did not correct inconsistencies between the AR roster and the health assessment, but the PTRACK file contains a best-guess birthdate variable. If the respondent knew the year of birth but not the month or day, this variable shows month and day as 98.

**Sex**. After the initial round, a respondent's sex was preprinted in the AR roster and collected on the cover of books. In cases of inconsistency between the roster and book covers, we undertook extensive checks on name, STAR information, and other data to ascertain a best guess for sex. The best-guess sex values are recorded in PTRACK.

**Marital Status**. Marital status was noted in the AR roster and on the covers of book 3. Various interviewer checks within the individual books required using marital status information from the book cover. In cleaning the data, we tried to make sure that marital status in the roster matched marital status on the book covers. We did not clean interviewer checks because that would have required complicated adjustments to skip patterns.

**Education Level**. The AR roster reported the highest level of schooling attained and the highest class completed within that level (AR30 and AR31), based on the answer provided by the respondent to Book K. Many respondents provide the information for themselves in response to questions in Books 3 or 5. Generally we recommend using the data provided by the individual rather than by a proxy if there is a choice.

**Earnings and Non-labor Income**. Module TK asked in depth about employment and labor earnings. The proxy book also addressed these topics. As insurance in case neither module was completed for some household members, we also included a question on earnings in the AR roster (AR37 and AR38). The existence of the AR data means that a measure of total household labor income can be computed, even if not all household members provided a book 3 or proxy book. However, data from TK are preferred because they come from the respondent or a knowledgeable proxy. TK data are likely to be more accurate also because earnings were addressed in the context of related questions.

**Parents' Survival Status**. The AR roster recorded PID numbers for each individual's mother and father (AR20 and AR24). If the mother or father was not a member of the household, codes were used to designate whether the parent was alive and living in another household or dead, and if dead, whether that death was attributable to the tsunami. Book 5, module BAA (parental information) explicitly asked the child respondent about each parent's survival status. The BAA data are preferred. Starting in STAR2, the Book 3s module BA asked those parent survival questions of adult respondents, and BA data are preferred for them.

**Timing of Marriage and Pregnancy.** The KF module provides data on marital timing and limited information on a woman's pregnancies. KF also provides some information on contraception. In rounds after STAR1, full pregnancy histories were administered in Book 4, and the Book 3 KF section was replaced with sections BR (reproduction) and KW (marriage)

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

## Parents, Children, and Spouses Identified in the AR Roster
The AR roster provides much information on relationships among current household members, as shown in the table below:

**Table 1**

| Variable | Information | Remarks |
|---|---|---|
| AR07 | Which member was designated household head and how other household members were related to that person | Sometimes this information indicates how members other than the household head were related. For example, if persons 3 and 4 were both children of the head, they were either full or half-siblings. If person 4 was the mother of the head and person 3 was the child of the head, person 4 was almost certainly the grandmother of person 3. In other cases the information is not definitive. For example, if persons 5 and 6 were both grandchildren of the head, they were likely to be siblings or cousins, but we do not know which from AR07. |
| AR12, AR20, and AR24 | PID numbers of an individual's spouse, birth mother, and birth father. | To find the education level of a child's parents, use the line numbers in AR20 and AR24 to link a child to its parents and thus to parents' education data either in the AR roster or in their individual book 3. |
| | | If a person's mother or father, or spouse is not listed on the household roster (because they were not a household member) there is no PID. |

Note one caution in using the AR data on family relationships. The accuracy of codes for AR21 and AR25 is not clear. The person completing the roster may not have known whether the father or mother was living or dead, especially in the displacement and confusion following the earthquake and tsunami in 2004. For the survival status of parents of children under 15 years of age in STAR, book 5 module BAA (parental information), is the preferred source of information because an explicit question was posed directly to the respondent. After STAR1, a BA module was also included in Book 3s, for respondents 15 and older.

**Classifying Relatives**

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

**AR07 vs. AR20/24.** Occasionally AR07 classified someone as the child of the head, but AR20 or AR24 did not list the head as the person's biological parent. The reason may be that

AR20/24 asked specifically about the biological parent, whereas AR07 asked more generally about the relationship to the head. Likewise, AR07 sometimes listed an individual as the parent of the household head, but that person's PID did not appear in the head's response for AR20/24 as a biological parent of the head.

**Divorce.** At the time of STAR1 some respondents' marriages had ended in divorce. During the "other" cleaning process we found responses indicating that someone was an ex-spouse or related to an ex-spouse. We created two new categories (ex-spouse and relative of ex-spouse) to account for this.

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

## Data Availability for Households and Individuals: HTRACK and PTRACK
Files named HTRACK and PTRACK indicate what data are available for households and respondents, respectively, in each survey wave.

### HTRACK
HTRACK also contains a record for all respondent households that were interviewed in STAR1 and, in subsequent waves includes households newly interviewed in the given wave (e.g., split-off households). HTRACK provides information on whether the household was interviewed in the given wave and, if so, whether data from books K, 1, and 2 are available.

HTRACK also contains household survey weights for all respondent households and an indicator for whether the cluster is rural.

Finally, HTRACK contains a measure of damage from the tsunami (the variable name is "damage"). The variable is defined at the level of the cluster. The measure takes on 3 values (no damage, some damage, and heavy damage), and is based on a combination of information collected from satellite imagery, survey supervisors, and community leaders.

### PTRACK
PTRACK contains a record for every individual who was interviewed in STAR1 and, in subsequent waves includes newly interviewed individuals in a given wave with information on interview status, sex, age in each wave, and variables that indicate our best effort to establish year, month, and date of birth.

PTRACK also provides individual survey weights for all STAR respondents.

**Table 2: STAR Household Survey Questionnaires**

| Respondent | Module | | Remarks |
|---|---|---|---|
| **Book K: Control Book** | | | |
| Interviewer and household head, spouse, or knowledgeable other person | AR | Household roster | Preprinted with the names of all targeted household members. |
| | CP | *See Note at end of table.* | |
| **Book 1: Expenditures, Prices, Housing Characteristics, Transfers** | | | |
| Wife of household head, household head, or other knowledgeable person | KS | Consumption | |
| | KH | Prices | |
| | KR | Housing characteristics | |
| | TR | Transfers | |
| | CP | *See Note at end of table.* | |
| **Book 2: Household Economy** | | | |
| Household head, wife of household head, or other household member | UT | Farm business | |
| | NT | Non-farm business | |
| | HU | Business assets | |
| | HR | Household assets | |
| | HI | Income | |
| | BT | Borrowing | |
| | HE | Health expenses (dropped after STAR1) | |
| | CP | *See Note at end of table.* | |
| **Book 3: Adult Individual Book** | | | |
| Each household member age 15 and older | DL | Education | |
| | KW | Marriage history (added in STAR2) | |
| | TK | Employment | |
| | HR | Individual assets | |
| | MA | Morbidities | |
| | KK | Health status | |

| | | |
|---|---|---|
| PB | Usual practices | |
| RJ | Outpatient utilization | |
| DK | Health diagnoses | |
| KF | Marriage and fertility (dropped after STAR1) | |
| KP | Psychological health | |
| SI | Attitudes | |
| GEI | Individual economic shocks | |
| MG | Migration | |
| BR | Reproduction (added in STAR2) | |
| SN | Social and kinship networks | |
| EX | Exposure | |
| PM | Community participation | |
| CX | Contraception (added in STAR2) | |
| CP | *See Note at end of table.* | |

**Book 3s:  Adult Individual Supplementary Book (begins in STAR2)**

| Each household member age 15 and older | BA | Non-coresident family and transfers |
|---|---|---|
| | KPB | Psychological health B |
| | CP | *See Note at end of table.* |

**Book 3b: Adult Individual Book for New Respondents (begins in STAR3)**

| New Household Member (AR01AB = 5) age 15 years or older | DL | Education |
|---|---|---|
| | KW | Marital History |
| | TK | Employment |
| | HR | Individual assets |
| | MA | Morbidities |
| | PB | Usual Practices |

| | KK | Health status |
|---|---|---|
| | RJ | Outpatient utilization |
| | DK | Health diagnoses |
| | KP | Psychological health |
| | KPB | Psychological health B |
| | SI | Attitudes |
| | GEI | Individual economic shocks |
| | MG | Migration |
| | BR | Reproduction |
| | SN | Social and kinship networks |
| | EX | Exposure |
| | PM | Community participation |
| | CP | *See Note at end of table.* |

| **Book 4: Book for Ever-married Women of Reproductive Age (begins in STAR2)** | | |
|---|---|---|
| Married or ever married woman age 15-49 | BR | Reproduction |
| | CH | Pregnancy History |
| | CX | Contraception (married only) |
| | CP | *See Note at end of table.* |

| **Book 5: Child Individual Book** | | |
|---|---|---|
| Each child, age 0–14 | DLA | Education |
| (usually answered by the mother if the child was less than 11 years old) | TKA | Employment |
| | TAA | Time allocation |
| | MAA | Morbidities |
| | KKA | Health status |
| | RJA | Outpatient utilization |
| | EXA | Exposure |
| | KPA | Psychological health |
| | KPAB | Child Health (psychol) B |

| | | |
|---|---|---|
| | MGA | Migration history |
| | BAA | Parental information |
| | SNA | Social and kinship networks |
| | INA | Institutional residence (added after STAR2) |
| | CP | *See Note at end of table.* |
| **Book 3 Proxy** | | |
| Someone who answered for the intended respondent to book 3. | Shortened version of other modules: Book 3: DL, TK, MA, KK, PB, DK, MG, SN Added in STAR2: BR (proxys of ever-marrieds) | |
| | CP | *See Note at end of table.* |
| **Book US Health Measurements** | | |
| Each household member | US | <u>Includes height, weight, blood pressure, hemoglobin</u> |
| | CP | *See Note at end of table.* |

**Note:** The CP module at the end of nearly every book asked the interviewer to record the conditions of the interview (who else was present, whether others provided assistance in answering questions), the respondent's level of attention, and any other relevant information about the interview environment. The interviewer could also add information to explain or clarify the respondent's answers.

## Appendix 1:
### Description of the STAR Questionnaire

This appendix expands on the summary presented in Section 2 for those interested in more detail about the survey instrument. While there is a great similarity across waves, users must be aware that some questions do change in order to reflect the passage of time since the tsunami and to reflect information learned in earlier waves. Users must **always** directly compare the questionnaires across waves when working with data from multiple waves to ensure they understand and adjust for any changes that may have occurred.

**Book K: Control Book and Household Roster**

The interviewer completed this book, or a portion of it, for all households targeted in STAR1. The cover of the book indicates (for each household, identified by HHID) whether the household was interviewed in the survey and if not, why not.

The AR Module (the household roster) lists all household members. After STAR1 the roster was preprinted with the name and characteristics of each household member from the previous wave. The roster was used to indicate household composition at the time of interview, and to enter basic information on age, sex, marital status, relationship to the head of the household, presence in the household of the individual's mother, father, and spouse, religion, whether the respondent worked or was in school, earnings in the last year, and highest level of education. Methods were intended to be sufficiently flexible to account for changes in household composition at each subsequent STAR resurvey. For individuals who had left the household between survey waves, information was collected on the reason for and date of departure and the person's current location. For individuals who joined the household since the previous survey, information was collected on the reason for and date of entry into the household.

**Book 1: Household Expenditures and Characteristics**

This book was answered by the household head, the spouse of the household head or by another person 18 years or older who is knowledgeable about household affairs. Module KS recorded information on expenditures for a variety of food and nonfood goods and services, including foods purchased and self-produced in the last week, personal care and household items bought during the last month, and durable goods bought in the last year. Module KH collected information about quantities and purchase prices for several staples. Module KR recorded information about ownership status and value of house and land, various household characteristics, and availability of and distance to certain public facilities. Module TR collected information about financial transfers, such as money for education or reconstruction, and in-kind transfers, such as food and building materials, into the household from different sources, including the government, NGOs, religious groups, family members, and friends and neighbors.

**Book 2: Household Economy**

Book 2 was answered by the household head, the spouse of the household head or by another person 18 years or older who is knowledgeable about household affairs. Modules UT and NT focused on household revenue, expenses, and value of household-owned agricultural and nonagricultural businesses. Module HU asked about the current value of household business assets (such as land, livestock, buildings, machinery, and transportation) as well as the value of any assets that were destroyed, sold, or given away. Module HR asked about the current value of household nonbusiness assets (e.g., land, livestock, jewelry), as well as asset ownership and ownership shares. Module HI asked about household-level non-labor income, by source. Module BT asked about money borrowed, received as a grant, or money that has been loaned out from the household.

**Book 3: Adult Individual Book**

This book elicited current and retrospective information from each household member age 15 and older.

**Education history.** Module DL recorded the highest level of education attended and highest grade completed, as well as basic literacy information. Information including the name and location of the most recent school attended was collected. Respondents were also asked about the highest level of schooling they expected to achieve.

**Marriage history.** Module KW, begun in STAR2, collected a detailed marital history starting with the current/most recent marriage and working backwards. The module asked for beginning and end dates for the marriage, the status of the marriage, and completed education of the spouse of that marriage. For current marriages, respondents were asked about assistance given to and received from a non-coresident spouse.

**Employment.** Module TK asked about respondents' current employment. Employment was categorized according to industry and position within the job, and information was recorded on hours worked and earnings. Unemployment information was collected for individuals without work.

**Individual non-labor income and assets.** Module HR collected information on individual-level economic well-being. Respondents were asked about the current value of their nonbusiness assets (e.g., land, livestock, jewelry), as well as asset ownership and ownership shares.

**Health status, physical performance, and usual practices.** Module KK asked about general health status and recent health history and physical functioning. Module MA asked about morbidities in the past four weeks and about experience with conditions symptomatic of heart disease, diabetes, and high blood pressure. Module PB asked about smoking and consumption of fruits and vegetables.

**Outpatient utilization and health diagnoses.** Module RJ collected information on outpatient visits during the last four weeks, including name and type of provider. Module DK recorded data on injuries and accidents in the past year, and more recent obvious illnesses.

**Marriage and fertility.** Module KF collected information about age and date of first marriage and number of children. Unmarried respondents were asked expectations of marriage timing and number of children. Married female respondents were asked about contraceptive use. This module was replaced in STAR2 with new Book 3 modules BR and KW along with Book 4 for ever-married women.

**Psychological health and attitudes.** Module KP asked respondents about symptoms of depression and post-traumatic stress response. Respondents were also asked about participation in post-disaster counseling and the usefulness thereof.

**Risk and Time Preferences**. Module SI asked respondents about time preferences (choosing between hypothetical payments sooner or later) and risk preferences (choosing between hypothetical certain and uncertain payouts).

**Individual economic shocks.** Module GEI asked respondents how life changed since the tsunami, as well as expectations for how long it will take for things to return to normal. Respondents also provided their opinions on the relative quality of their lives on a 6-step ladder, before the tsunami, at the time of the interview, and five years in the future.

**Migration history.** Module MG collected information on the geographic mobility of individuals, as well as the causes and consequences of migratory movements, including short-stay and circulatory migration. Information was recorded about the respondent's location at birth, age 12, and each

subsequent location where a move crossed a *desa* (village) boundary and lasted for 6 months or longer. For each move, data were collected on dates and locations, motivation for moving, and distance moved.

**Reproduction.** Module BR, added in STAR2, asked ever-married non-Book 4 respondents about the number of children of each sex  they had given birth to/fathered, both currently living, split by co-resident and non-co-resident status, and deceased. For the deceased, respondents were asked for the number that passed away at the time of the tsunami, if any, besides the total number. Never married male and female respondents were only asked questions about fertility preferences, which were also asked of the others.  Ever-married women age 15-49 skipped the BR module in Book 3 as they would be asked these questions in Book 4.

**Social and kinship network***.* Module SN asked respondents about the number of family members and friends and neighbors that could be called upon for certain help or support. Specific questions were asked about the survival of critical family members.  For STAR2 onward, specific questions on the survival status of parents was moved to Book 3S module BA.

**Exposure***.* Module EX recorded data on measures of exposure to the earthquake and tsunami, such as feeling the earthquake, hearing people shouting, physically being swept away by the wave, and living afterward in destroyed areas.

**Community participation.** Module PM considered community development activities, which have long been a backbone of development in Indonesia. Respondents were asked about participation in, contributions of time and money to, and perceived benefits from, a slate of community development and reconstruction activities. Questions were included on participation in rotating credit schemes (*arisan*) and knowledge and use of credit sources.

### Book 3S: Book 3 Supplement
This book was administered in waves after STAR1 to collect information from Book 3 respondents for two additional modules. Proxy respondents were not given any portion Book 3S.

**Non-Coresident Family and Transfers.** Module BA collected information on non-coresident and deceased parents including transfers; information was collected about brothers and sisters in total and transfers between them and the respondent. Module BA also contains a detailed roster of non-coresident biological children plus those that died in the last 10 years which also contains information on transfers.

**Adult Mental Health.**  Module KPB covered an additional set of psychological health questions beyond those in Section KP in Book 3 and included questions related to the CES-D depression scale and Post-Traumatic Stress Disorder Checklist (PCL-C).

### Book 3b: First time adult respondent retrospectives
This book was administered in waves after STAR2 to collect retrospective information for first-time adult respondents, without repeating life histories for panel respondents.  The modules are mainly the same as those in Book 3.

### Book 4: Information for Ever-Married Women
This book was administered in waves after STAR1 to collect information on children ever born, a full pregnancy history, and knowledge and use of contraceptives.

**Book 5: Child Information**

This book was administered to household members younger than 15. For children younger than 11, the mother, female guardian, or household caretaker answered the questions. Children between the ages of 11 and 14 were allowed to respond for themselves if they wished. Generally, each module paralleled a module in the adult questionnaire (Book 3), with age-appropriate modifications. Modules that paralleled the adult questionnaire included: child's educational status and schooling disruptions (module DLA), employment status (module TKA), time allocation (module TAA), morbidities and general health characteristics (modules MAA and KKA), outpatient utilization (module RJA), exposure to the earthquake and tsunami (module EXA), child psychological health (module KPA), migration history (module MGA), social and kinship network (module SNA). Paralleling the more detailed adult versions in Book 3s, Book 5 included module BAA (first done in STAR1 for child respondents), which collected information on the respondent's parents' survival status, location if non-coresident, and inheritance if both deceased, plus information on other surviving kin, and module KPAB (added in STAR2), which contained additional questions beyond those in Module KPA related to depression and Post-Traumatic Stress Disorder. Module INA (added after STAR2), and unique to Book 5, recorded data on type of institution in cases when the respondent lived in an institution at the time of interview.

**Book US: Health Measurements**

This book was administered in waves after STAR1 to all household members, with some tests being age-dependent.  All ages were measured for height and weight, those over 40 additionally had waist and hip circumferences measured. Those age 15 and over had blood pressure and pulse readings taken. Those age 1 and older had hemoglobin measured and an SPRT blood sample taken. Respondents were also asked whether they were fasting that day, if they were currently taking medication for anemia, blood pressure or diabetes, and if they had taken iron pills within the last 4 months.

## Appendix 2:
## Description of the STAR2 Data

### 1. Introduction

This appendix focuses on the second round of STAR (STAR2). STAR2 was conducted between January, 2006 and May, 2007. STAR2 targeted 7,160 households first interviewed in STAR1 and an additional 1,228 households that were originally selected for STAR1 but that could not be interviewed during the STAR1 field period because of security risks. These households were interviewed for the first time in STAR2 and are identified by the variable hhadd02 in the htrack file. STAR2 also includes 1,403 "split-off" households formed when an individual from a STAR1 household moves and is tracked and interviewed in a new household (for split-offs, the household identifier, idhh02, contains a "2" in the 8$^{th}$ digit, denoting that the household first appears in STAR2). STAR2 provides information about a total of 9,791 households.  Outcomes for STAR2 households are depicted in Figure 1.

FIGURE 1
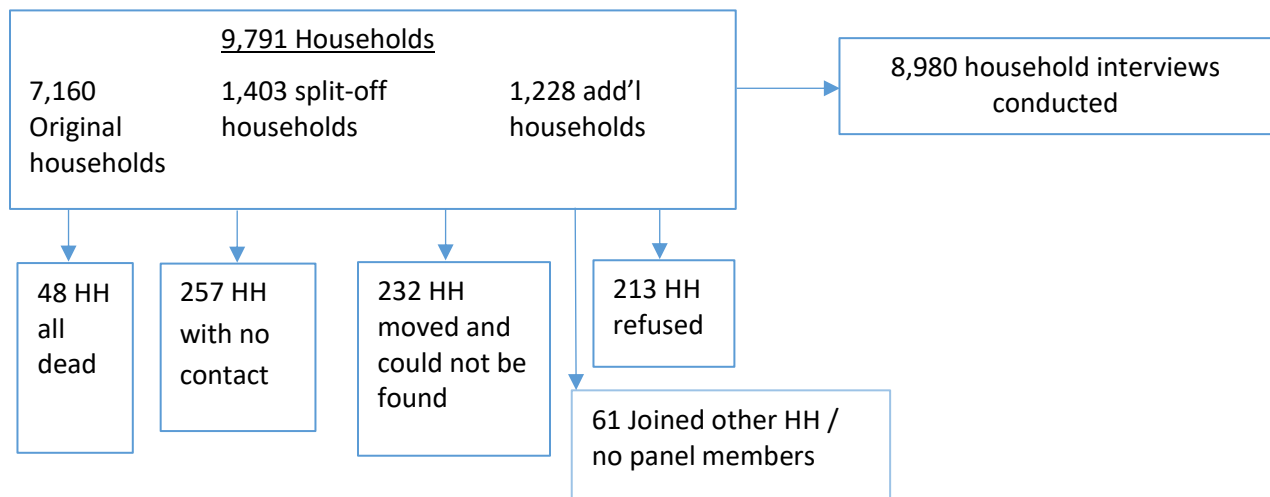


Turning to individuals, there were 38,466 people in either a STAR1 or STAR2 household. A total of 200 individuals died between STAR1 and STAR2. Among 38,266 survivors at STAR2, 36,223 (94.7%) were in interviewed households. Another 1,464 (3.8%) were part of households that were not reinterviewed in STAR2, and 579 (1.5%) moved out of a STAR1 household that was interviewed in STAR2, and could not be tracked to their new destination. In STAR2 individual Book 3 or Book 5 interviews were conducted with 34,907 of those in interviewed households.

With respect to attrition, of the 28,206 STAR1 sample members who survived to STAR2, 94% were followed up in STAR2 and located in a household that was interviewed in STAR2, and we  completed individual interviews with over 90%.

### 2. Survey Instruments

Books that were included in STAR1 (Books K, 1, 2, 3, and 5) are repeated in STAR2. Some books contain new modules. For example, Book 3 of STAR2 contains sections KW and BR, which collect more detailed information about marriages and children ever born than was available in the KF section of Book 3 in STAR1.

In addition, STAR2 includes three new books: 3S, 4, and US. Book 3S, addressed to individuals age 15 and older, contains detailed questions on family networks, including relationships with parents, siblings, and children. A more detailed battery of questions on post-traumatic stress and depression is included as well. Book 4, addressed to ever-married women between the ages of 15 and 49, asks questions on children ever born, collects a full pregnancy history, and asks a number of limited information on contraceptive use. Book US covers data from physical health assessments. These include each household member's height, weight, blood pressure, and hemoglobin status, among other measures.

## 3. STAR File Structure and Naming Conventions
### Identifiers and Level of Observation
Identifiers are discussed in detail in the documentation for STAR1 (pp 4-5).

In STAR2 the unique identifiers are IDHH02, IDP02, and IDPLINK. Briefly, IDHH02 is a unique 9 digit identifier assigned to each household. The first three digits correspond to the survey cluster. Digits 4-6 identify the origin household (the household originally targeted for an interview in STAR1). Digits 7 and 8 denote the round of STAR in which the household originated. These digits are "01" for all households targeted for interview in STAR1. As noted above, the digits "02" signify that the household is a split-off household identified in STAR2. The final digit designates the order in which a household is found in a particular wave (it is 0 for all origin households, and greater than 0 for all split-off households).

The person identifier is IDP02. It is the line number of that individual in the household roster in STAR2.

IDPLINK identifies individuals across households and across rounds of the STAR data. IDPLINK is formed by combining the IDHH and IDP from the household and roster number in which the individual was first listed. IDPLINK is constant across waves and across whatever households an individual is currently or has ever been a member.

When merging across books within STAR2, it is appropriate to merge on IDHH02 and IDP02. When merging across waves (i.e. STAR1 and STAR2), it is appropriate to merge on IDPLINK.

As in STAR1, in some data files the level of observation is something other than a household or individual. Usually the variable that defines observations in these files is named XXTYPE, where XX identifies the associated module (see pages 5 and 9 in the STAR documentation).

For Section US (health measurements) the preprinted household roster with updates on current household composition served as the base for obtaining US measurements. For this reason, some individuals have an IDPLINK that appears twice in US (although as with the AR roster, idhh02 and idp02 will appear only once).

### Question Numbers and Variable Names
Within modules, question numbers and variable names in STAR2 match STAR1 when the questions are the same. New or different questions are given different numbers and variable names.

## 4. Special Features of the STAR Data
**Household Roster**: The household roster (AR) in follow-up rounds of STAR is administered as a preprinted list of all household members ever belonging to the household. Each member's status in the household is updated, indicating whether they are still residing in the household, have moved out, or

have died. In many instances members who have moved out are interviewed in a new split-off household. This protocol means that an individual may appear in the roster of more than one household (though he or she will be identified as a *current member* of only one household in any round). In the AR data files, this protocol generates duplicate observations on IDPLINK (the permanent identifier carries across rounds and households), but unique observations on IDHH02 and IDP02 (and in all rounds other than STAR1), because these identifiers are specific to a particular round and household. Note that the variable AR02 serves to identify whether an individual is a current member of the household (in which case AR02=1).

**HTRACK and PTRACK**: These files are designed to provide an accounting system for households and individuals across multiple rounds of the STAR data. HTRACK is a household-level data file that includes a record for each household that was a target in STAR1 or STAR2 and a record for each split-off household identified in STAR2. The file contains household IDs in each round, results codes for each round, the interview date (randomized for privacy protection, as described on p 9 of the STAR documentation). In additional, household-level weights, raked to the projected population of the STAR districts in 2006, are included. PTRACK in an individual-level file that contains identifiers for each survey round, demographic and interview information for STAR2, and individual-level weights (also raked to the projected 2006 population by age, sex, district, and urban-rural status). To facilitate using the data, HTRACK and PTRACK for STAR2 carry over a small amount of information from STAR1. PTRACK contains a measure, agebg_02, that represents our best guess of a person's age at interview (based on comparisons of information across multiple sources). There is a also a variable "sex" that represents what we believe a person's biological sex to be. Note that in some cases agebg_02 or sex may differ from the value recorded in the household roster or the book cover.

# Appendix 3:
# Description of the STAR3 Data

## 1. Introduction

This appendix focuses on the third round of STAR (STAR3). STAR3 was conducted between July, 2007 and April, 2008. STAR3 targeted all STAR1 and STAR2 households.  Outcomes in STAR3 for STAR2 households are depicted in Figure 1.

FIGURE 1
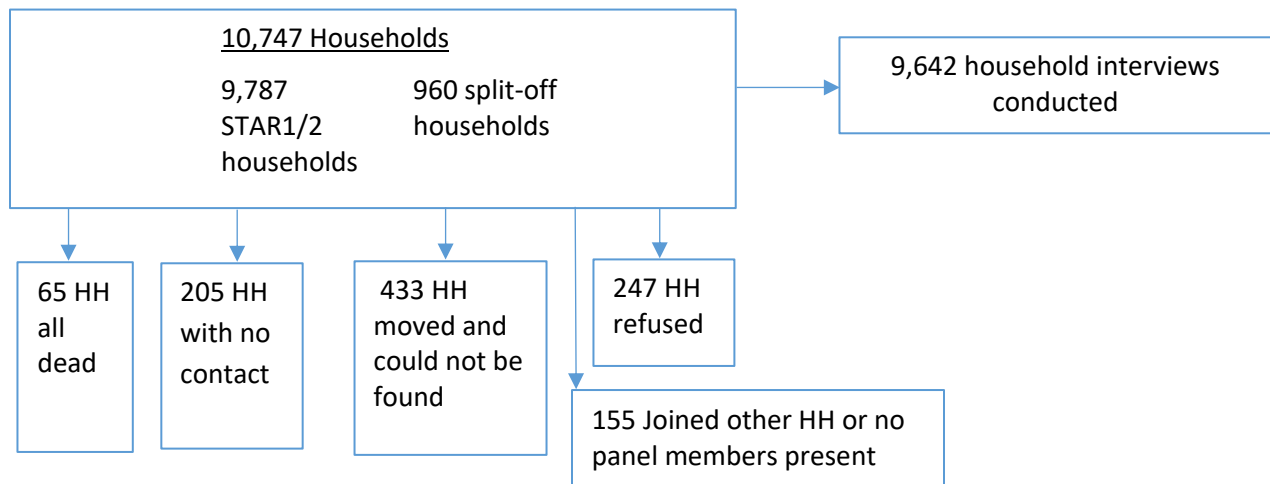


Turning to individuals, 38,456 individuals were listed in STAR1 or STAR2 households. A total of 200 of these individuals died between STAR1 and STAR2. Among survivors at STAR3, 34,468 (90%) were in interviewed households. Another 1086 (2.8%) were part of households that were not reinterviewed in STAR3, and 1,368 (3.6%) moved out of a household that was interviewed in STAR3 but could not be tracked and interviewed in their new destination. A total of 1,071 individuals (2.8%) refused to be interviewed. In STAR3 there were 2,591 new members of STAR households (these individuals were not part of STAR1 or STAR2 households).  In STAR3 individual Book 3 or Book 5 interviews were conducted with 36,110 of those in interviewed households.

## 2. Survey Instruments

The basic structures of the books in STAR1 and STAR2 are repeated in STAR3. In addition, STAR3 introduces Book 3B and modifies Book 4. Book 3B and Book 4 include the histories first fielded in STAR2. They are administered to new respondents who did not provide history information in STAR2. Accompanying the introduction of Books 3B and 4, Book 3 no longer collects full histories, since these were collected in STAR2 and do not need to be repeated. Book 4 for panel respondents was eliminated, for the same reason. Instead, women ages 15-49 who are panel respondents answer sections CH and CX as part of Book 3. In most instances questions in Books 3 and 3B are the same when they have the same numbers and variable names. However, this is not true for the KW module. In STAR3, Book 3 for panel respondents updates marital status and collects data about any changes since 2006. Book 3B asks a full marriage history. In STAR3, the KPAB section is part of Books 3 and 3B rather than Book 3S. Book 3S contains the BA module. Book 5 drops the MGA and KPA sections.

## 3. STAR File Structure and Naming Conventions
### Identifiers and Level of Observation
Identifiers are discussed in detail in the documentation for STAR1 (pp 4-5).

In STAR3 the unique identifiers are IDHH03, IDP03, and IDPLINK. They are similar to the unique identifiers in STAR2.  Briefly, IDHH03 is a unique 9 digit identifier assigned to each household. The first three digits correspond to the survey cluster. Digits 4-6 identify the origin household (the household originally targeted for an interview in STAR1). Digits 7 and 8 denote the round of STAR in which the household originated. These digits are "01" for all households targeted for interview in STAR1. As noted above, the digits "03" signify that the household is a split-off household identified in STAR3. The final digit designates the order in which a household is found in a particular wave (it is 0 for all origin households, and greater than 0 for all split-off households).

The person identifier is IDP03. It is the line number of that individual in the household roster in STAR3.

IDPLINK identifies individuals across households and across rounds of the STAR data. IDPLINK is formed by combining the IDHH and IDP from the household and roster number in which the individual was first listed. IDPLINK is constant across waves and across whatever households an individual is currently or has ever been a member.

When merging across books within STAR3, it is appropriate to merge on IDHH03 and IDP03. When merging across waves (i.e. STAR1-STAR3), it is appropriate to merge on IDPLINK.

As in STAR1 and STAR2, in some data files the level of observation is something other than a household or individual. Usually the variable that defines observations in these files is named XXTYPE, where XX identifies the associated module (see pages 5 and 9 in the STAR documentation).

For Section US (health measurements) the preprinted household roster with updates on current household composition served as the base for obtaining US measurements. For this reason, some individuals have an IDPLINK that appears twice in US (although as with the AR roster, idhh03 and idp03 will appear only once).

### Question Numbers and Variable Names
Within modules, question numbers and variable names in STAR3 match earlier waves when the questions are the same. New or different questions are given different numbers and variable names. An exception is the KW (marriage) section of Book 3 and 3B. Question numbers are the same but the content of the questions differ, with a full marriage history collected for new respondents answering Book 3B, whereas only an update is collected from panel respondents answering Book 3B.

## 4. Special Features of the STAR Data
**Household Roster**:  The household roster (AR) in follow-up rounds of STAR is administered as a preprinted list of all household members ever belonging to the household. Each member's status in the household is updated, indicating whether they are still residing in the household, have moved out, or have died. In many instances members who have moved out are interviewed in a new split-off household. This protocol means that an individual may appear in the roster of more than one household (though he or she will be identified as a *current member* of only one household in any round). In the AR data files, this protocol generates duplicate observations on IDPLINK (the permanent identifier carries across rounds and households), but unique observations on IDHH03 and IDP03 (and in all rounds other

than STAR1), because these identifiers are specific to a particular round and household. Note that the variable AR02 serves to identify whether an individual is a current member of the household (in which case AR02=1).

**HTRACK and PTRACK**: These files are designed to provide an accounting system for households and individuals across multiple rounds of the STAR data. HTRACK is a household-level data file that includes a record for each household that was a target in STAR1, STAR2, or STAR3 and a record for each split-off household identified in STAR2 or STAR3.  The file contains household IDs in each round, results codes for each round, the interview date (randomized for privacy protection, as described on p 9 of the STAR documentation). In additional, household-level weights, raked to the projected population of the STAR districts in 2007, are included. PTRACK in an individual-level file that contains identifiers for each survey round, demographic and interview information for STAR2, and individual-level weights (also raked to the projected 2007 population by age, sex, district, and urban-rural status). To facilitate using the data, HTRACK and PTRACK for STAR3 carry over a small amount of information from STAR1.  PTRACK contains a measure, agebg_03, that represents our best guess of a person's age at interview (based on comparisons of information across multiple sources). There is also a variable "sex" that represents what we believe a person's biological sex to be. Note that in some cases agebg_03 or sex may differ from the value recorded in the household roster or the book cover.

## Appendix 4:
## Description of the STAR4 Data

### 1. Introduction

This appendix focuses on the fourth round of STAR (STAR4). STAR4 was conducted between July, 2007 and April, 2008. STAR4 targeted all 10,748 STAR1, STAR2, and STAR3 households as well as 1,186 new households formed by panel respondents creating new households. Outcomes in STAR4 for STAR3 households are depicted in Figure 1.

FIGURE 1
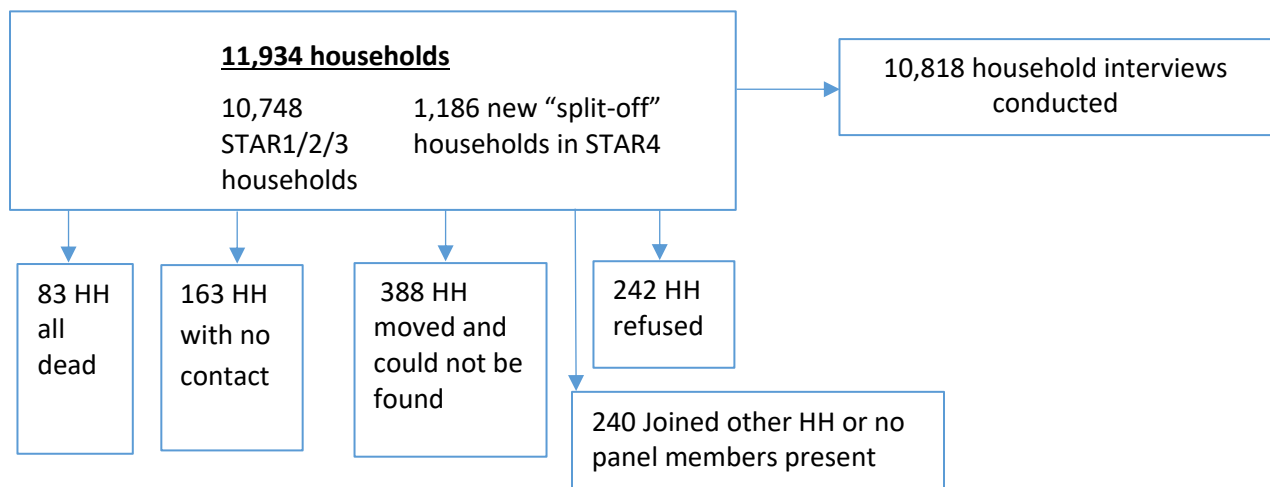


Turning to individuals, 41,054 individuals were listed in STAR1, STAR2, or STAR3 households. A total of 741 of these individuals died between STAR1 and STAR4. Among survivors at STAR4, 38,741 (91%) were in interviewed households. Another 885 (2%) were part of households that were not reinterviewed in STAR4, and 2,067 (4.8%) moved out of a household that was interviewed in STAR4 but could not be tracked and interviewed in their new destination. A total of 988 individuals (2.3%) refused to be interviewed. In STAR4 there were 2,371 new members of STAR households (these individuals were not part of STAR1, STAR2, or STAR3 households). In STAR4 individual Book 3 or Book 5 interviews were conducted with 37,719 of those in interviewed households.

### 2. Survey Instruments

The basic structures of the books in STAR4 are similar to the books in STAR3. In addition, STAR4 introduces two new modules in Books 3 and 3B: CO (word recall), BL (business loans), and a new book, EK, which consists of 16 items from the Philippines Nonverbal Intelligence Test and is administered to children age 7-14. In STAR4, as in STAR3, Book 3B and Book 4 include the histories first fielded in STAR2. They are administered to new respondents who did not provide history information in STAR3. In most instances questions in Books 3 and 3B are the same when they have the same numbers and variable names. However, this is not true for the KW module. Book 3 for panel respondents updates marital status and collects data about recent changes, whereas Book 3B asks a full marriage history.

### 3. STAR File Structure and Naming Conventions
**Identifiers and Level of Observation**

28

Identifiers are discussed in detail in the documentation for STAR1 (pp 4-5).

In STAR4 the unique identifiers are IDHH04, IDP04, and IDPLINK. They are similar to the unique identifiers in previous waves.  Briefly, IDHH04 is a unique 9 digit identifier assigned to each household. The first three digits correspond to the survey cluster. Digits 4-6 identify the origin household (the household originally targeted for an interview in STAR1). Digits 7 and 8 denote the round of STAR in which the household originated. These digits are "01" for all households targeted for interview in STAR1. As noted above, the digits "04" signify that the household is a split-off household identified in STAR4. The final digit designates the order in which a household is found in a particular wave (it is 0 for all origin households, and greater than 0 for all split-off households).

The person identifier is IDP04. It is the line number of that individual in the household roster in STAR4.

IDPLINK identifies individuals across households and across rounds of the STAR data. IDPLINK is formed by combining the IDHH and IDP from the household and roster number in which the individual was first listed. IDPLINK is constant across waves and across whatever households an individual is currently or has ever been a member.

When merging across books within STAR4, it is appropriate to merge on IDHH04 and IDP04. When merging across waves (i.e. STAR1-STAR4), it is appropriate to merge on IDPLINK.

As in previous waves, in some data files the level of observation is something other than a household or individual. Usually the variable that defines observations in these files is named XXTYPE, where XX identifies the associated module (see pages 5 and 9 in the STAR documentation).

For Section US (health measurements) the preprinted household roster with updates on current household composition served as the base for obtaining US measurements. For this reason, some individuals have an IDPLINK that appears twice in US (although as with the AR roster, idhh04 and idp04 will appear only once).

### Question Numbers and Variable Names
Within modules, question numbers and variable names in STAR4 match earlier waves when the questions are the same. New or different questions are given different numbers and variable names. An exception is the KW (marriage) section of Book 3 and 3B. Question numbers are the same but the content of the questions differs, with a full marriage history collected for new respondents answering Book 3B, whereas only an update is collected from panel respondents answering Book 3.

## 4. Special Features of the STAR Data
**Household Roster**:  The household roster (AR) in follow-up rounds of STAR is administered as a preprinted list of all household members ever belonging to the household. Each member's status in the household is updated, indicating whether they are still residing in the household, have moved out, or have died. In many instances members who have moved out are interviewed in a new split-off household. This protocol means that an individual may appear in the roster of more than one household (though he or she will be identified as a *current member* of only one household in any round). In the AR data files, this protocol generates duplicate observations on IDPLINK (the permanent identifier carries across rounds and households), but unique observations on IDHH03 and IDP03 (and in all rounds other than STAR1), because these identifiers are specific to a particular round and household. Note that the

variable AR02 serves to identify whether an individual is a current member of the household (in which case AR02=1).

**HTRACK and PTRACK**: These files are designed to provide an accounting system for households and individuals across multiple rounds of the STAR data. HTRACK is a household-level data file that includes a record for each household that was a target in STAR1, STAR2, STAR3, or STAR4 and a record for each split-off household identified in STAR2, STAR3, or STAR4.  The file contains household IDs in each round, results codes for each round, the interview date (randomized for privacy protection, as described on p 9 of the STAR documentation). In additional, household-level weights, raked to the projected population of the STAR districts in 2008, are included. PTRACK in an individual-level file that contains identifiers for each survey round, demographic and interview information, and individual-level weights (also raked to the projected 2008 population by age, sex, district, and urban-rural status). To facilitate using the data, HTRACK and PTRACK for STAR4 carry over a small amount of information from previous rounds of data. PTRACK contains a measure, agebg_04, that represents our best guess of a person's age at interview (based on comparisons of information across multiple sources). There is also a variable "sex" that represents what we believe a person's biological sex to be. Note that in some cases agebg_04 or sex may differ from the value recorded in the household roster or the book cover.

## Appendix 5:
## Description of the STAR5 Data

### 1. Introduction

This appendix focuses on the fifth round of STAR (STAR5). STAR5 was conducted between Nov, 2009 and Nov, 2010. STAR5 targeted all 11,934 STAR1-STAR4 households as well as 1,781 new households formed by panel respondents creating new households.  Outcomes in STAR5 for STAR4 households are depicted in Figure 1.

FIGURE 1
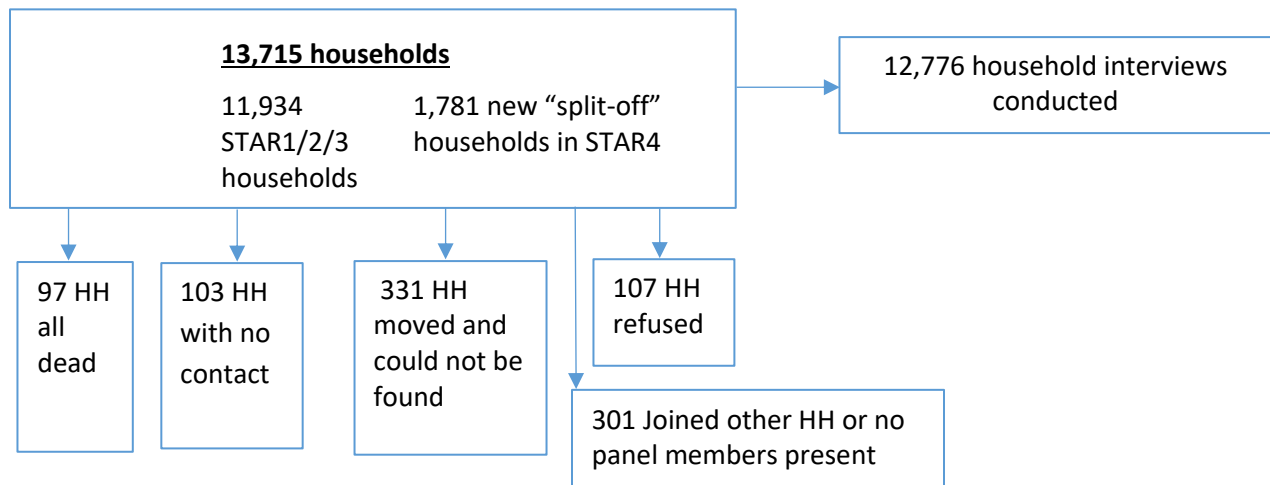


Turning to individuals, 43,389 individuals were listed in STAR1-STAR4 households. A total of 1,117 of these individuals died between STAR1 and STAR5. Among survivors at STAR5, 41,404 (92%) were in interviewed households. Another 728 (1.6%) were part of households that were not reinterviewed in STAR5, and 2,523 (5.6%) moved out of a household that was interviewed in STAR5 but could not be tracked and interviewd in their new destination. A total of 432 individuals (1%) refused to be interviewed. In STAR5 there were 2,815 new members of STAR households (these individuals were not part of STAR1, STAR2, STAR3 or STAR4 households).  In STAR5 individual Book 3 or Book 5 interviews were conducted with 41,327 of those in interviewed households.

### 2. Survey Instruments

The basic structures of the books in STAR5 are similar to the books in STAR4. In STAR5 Book 3B and Book 4 include the histories first fielded in STAR2. They are administered to new respondents who did not provide history information in STAR4. In most instances questions in Books 3 and 3B are the same when they have the same numbers and variable names. However, this is not true for the KW module. Book 3 for panel respondents updates marital status and collects data about recent changes, whereas Book 3B asks a full marriage history.

### 3. STAR File Structure and Naming Conventions
#### Identifiers and Level of Observation
Identifiers are discussed in detail in the documentation for STAR1 (pp 4-5).

In STAR5 the unique identifiers are IDHH05, IDP05, and IDPLINK. They are similar to the unique identifiers in previous waves.  Briefly, IDHH05 is a unique 9 digit identifier assigned to each household. The first three digits correspond to the survey cluster. Digits 4-6 identify the origin household (the household originally targeted for an interview in STAR1). Digits 7 and 8 denote the round of STAR in which the household originated. These digits are "01" for all households targeted for interview in STAR1. As noted above, the digits "05" signify that the household is a split-off household identified in STAR5. The final digit designates the order in which a household is found in a particular wave (it is 0 for all origin households, and greater than 0 for all split-off households).

The person identifier is IDP05. It is the line number of that individual in the household roster in STAR5.

IDPLINK identifies individuals across households and across rounds of the STAR data. IDPLINK is formed by combining the IDHH and IDP from the household and roster number in which the individual was first listed. IDPLINK is constant across waves and across whatever households an individual is currently or has ever been a member.

When merging across books within STAR5, it is appropriate to merge on IDHH05 and IDP05. When merging across waves (i.e. STAR1-STAR5), it is appropriate to merge on IDPLINK.

As in previous waves, in some data files the level of observation is something other than a household or individual. Usually the variable that defines observations in these files is named XXTYPE, where XX identifies the associated module (see pages 5 and 9 in the STAR documentation).

For Section US (health measurements) the preprinted household roster with updates on current household composition served as the base for obtaining US measurements. For this reason, some individuals have an IDPLINK that appears twice in US (although as with the AR roster, idhh04 and idp04 will appear only once).

### Question Numbers and Variable Names
Within modules, question numbers and variable names in STAR5 match earlier waves when the questions are the same. New or different questions are given different numbers and variable names. An exception is the KW (marriage) section of Book 3 and 3B. Question numbers are the same but the content of the questions differs, with a full marriage history collected for new respondents answering Book 3B, whereas only an update is collected from panel respondents answering Book 3.

## 4. Special Features of the STAR Data
**Household Roster**:  The household roster (AR) in follow-up rounds of STAR is administered as a preprinted list of all household members ever belonging to the household. Each member's status in the household is updated, indicating whether they are still residing in the household, have moved out, or have died. In many instances members who have moved out are interviewed in a new split-off household. This protocol means that an individual may appear in the roster of more than one household (though he or she will be identified as a *current member* of only one household in any round). In the AR data files, this protocol generates duplicate observations on IDPLINK (the permanent identifier carries across rounds and households), but unique observations on IDHH03 and IDP03 (and in all rounds other than STAR1), because these identifiers are specific to a particular round and household. Note that the variable AR02 serves to identify whether an individual is a current member of the household (in which case AR02=1).

**HTRACK and PTRACK**: These files are designed to provide an accounting system for households and individuals across multiple rounds of the STAR data. HTRACK is a household-level data file that includes a record for each household that was a target in STAR1, STAR2, STAR3, STAR4, or STAR5 and a record for each split-off household identified in STAR2, STAR3, STAR4 or STAR5.  The file contains household IDs in each round, results codes for each round, the interview date (randomized for privacy protection, as described on p 9 of the STAR documentation). In additional, household-level weights, raked to the projected population of the STAR districts in 2009, are included. PTRACK in an individual-level file that contains identifiers for each survey round, demographic and interview information, and individual-level weights (also raked to the projected 2009 population by age, sex, district, and urban-rural status). To facilitate using the data, HTRACK and PTRACK for STAR5 carry over a small amount of information from previous rounds of data.  PTRACK contains a measure, agebg_05, that represents our best guess of a person's age at interview (based on comparisons of information across multiple sources). There is also a variable "sex" that represents what we believe a person's biological sex to be. Note that in some cases agebg_05 or sex may differ from the value recorded in the household roster or the book cover.

# Appendix 6:
# Description of the STAR6 Data

## 1. Introduction

This appendix focuses on the sixth round of STAR (STAR6). STAR6 was conducted between September, 2014 and June, 2016. STAR6 targeted all 13,716 STAR1-STAR5 households as well as 3,916 new households formed by panel respondents creating new households.  Outcomes in STAR6 for STAR5 households are depicted in Figure 1.

FIGURE 1
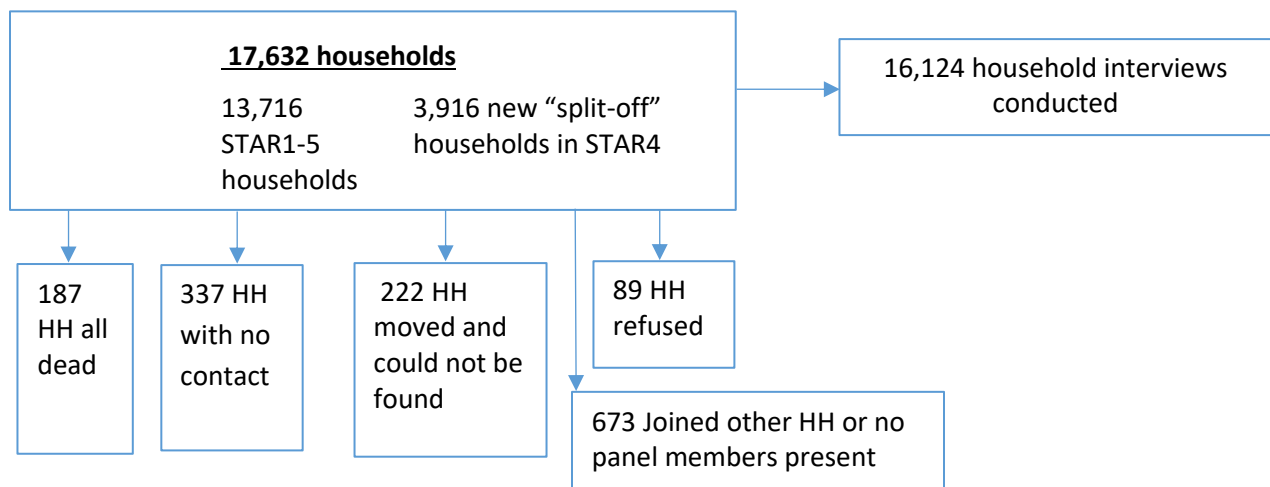


Turning to individuals, 46,160 individuals were listed in STAR1-STAR5 households, of whom 45,033 were alive and 1,123 were deceased. An additional 1,350 individuals died between STAR5 and STAR6. Between STAR5 and STAR6 10,007 individuals joined STAR households as new members. At STAR6, 50,027 individuals (93% of survivors) were in interviewed households. Another 626 (1.2%) were part of households that were not reinterviewed in STAR6, and 2,748 individuals (~5%) moved out of a household that was interviewed in STAR6 but could not themselves be tracked and interviewed in their new destination. A total of 339 individuals (0.6%) were in households that refused to be interviewed. In STAR6 there were 10,007 new members of STAR households (these individuals were not part of STAR1-STAR5 households).  In STAR6 individual Book 3 or Book 5 interviews were conducted with 49,710 of those in interviewed households.

## 2. Survey Instruments

The basic structures of the books in STAR6 are similar to the books in STAR5. In STAR6 Book 3B and Book 4 include the histories first fielded in STAR2. They are administered to new respondents who did not provide history information in STAR5. Respondents who provided histories in the past answer Book 3, which collects information on events that occurred since 2009 (for modules that collect "history" information, such as education, marriage, and migration, among others), or over the past 12 months (for example, receipt of inpatient care). In most instances questions in Books 3 and 3B that share the same numbers and variable names provide data about the same constructs. However, this is not true for the KW module. Book 3 for panel respondents updates marital status and collects data about recent changes, whereas Book 3B asks a full marriage history.

There are some changes and new additions to the questionnaire. Some sections have moved from Books 3 or 3B to Book 3S, to better balance the length of the books. For example, sections PM (Community Participation) and SI (Attitudes toward Risk) are now in Book 3S, which also contains several new sections. The new sections are PS (Pittsburgh Sleep), SS (Social Support), BL (Business Loan), and BI (Big 5 Personality Inventory). Because the addition of the new sections increases the length of the questionnaire to durations that are unappealing for new respondents (who also answer the full histories), new respondents answer a subset of the new modules based on a randomization scheme (for example, new respondents whose random letter is A answer the PS module on sleep, whereas those with other letters skip PS but answer other sections in Book 3S). Book 5 also includes the Big 5 Personality Inventory. A new EK module was adopted to test cognitive skills among individuals above age 15.

## 3. STAR File Structure and Naming Conventions
### Identifiers and Level of Observation
Identifiers are discussed in detail in the documentation for STAR1 (pp 4-5).

In STAR6 the unique identifiers are IDHH06, IDP06, and IDPLINK. They are similar to the unique identifiers in previous waves.  Briefly, IDHH06 is a unique 9 digit identifier assigned to each household. The first three digits correspond to the survey cluster. Digits 4-6 identify the origin household (the household originally targeted for an interview in STAR1). Digits 7 and 8 denote the round of STAR in which the household originated. These digits are "01" for all households targeted for interview in STAR1. As noted above, the digits "06" signify that the household is a split-off household identified in STAR6. The final digit designates the order in which a household is found in a particular wave (it is 0 for all origin households, and greater than 0 for all split-off households).

The person identifier is IDP06. It is the line number of that individual in the household roster in STAR6.

IDPLINK identifies individuals across households and across rounds of the STAR data. IDPLINK is formed by combining the IDHH and IDP from the household and roster number in which the individual was first listed. IDPLINK is constant across waves and across whatever households an individual is currently or has ever been a member.

When merging across books within STAR6, it is appropriate to merge on IDHH06 and IDP06. When merging across waves (i.e. STAR1-STAR6), it is appropriate to merge on IDPLINK.

As in previous waves, in some data files the level of observation is something other than a household or individual. Usually the variable that defines observations in these files is named XXTYPE, where XX identifies the associated module (see pages 5 and 9 in the STAR documentation).

For Section US (health measurements) the preprinted household roster with updates on current household composition served as the base for obtaining US measurements. For this reason, some individuals have an IDPLINK that appears twice in US (although as with the AR roster, idhh06 and idp06 will appear only once).

### Question Numbers and Variable Names
Within modules, question numbers and variable names in STAR6 match earlier waves when the questions are the same. New or different questions are given different numbers and variable names. An

exception is the KW (marriage) section of Book 3 and 3B. Question numbers are the same but the content of the questions differs, with a full marriage history collected for new respondents answering Book 3B, whereas only an update is collected from panel respondents answering Book 3.

## 4. Special Features of the STAR Data

**Household Roster**:  The household roster (AR) in follow-up rounds of STAR is administered as a preprinted list of all household members ever belonging to the household. Each member's status in the household is updated, indicating whether they are still residing in the household, have moved out, or have died. In many instances members who have moved out are interviewed in a new split-off household. This protocol means that an individual may appear in the roster of more than one household (though he or she will be identified as a *current member* of only one household in any round). In the AR data files, this protocol generates duplicate observations on IDPLINK (the permanent identifier carries across rounds and households), but unique observations on IDHH03 and IDP03 (and in all rounds other than STAR1), because these identifiers are specific to a particular round and household. Note that the variable AR02 serves to identify whether an individual is a current member of the household (in which case AR02=1).

**HTRACK and PTRACK**: These files are designed to provide an accounting system for households and individuals across multiple rounds of the STAR data. HTRACK is a household-level data file that includes a record for each household that was a target in STAR1-STAR5 or STAR6 and a record for each split-off household identified in STAR2-STAR6.  The file contains household IDs in each round, results codes for each round, the interview date (randomized for privacy protection, as described on p 9 of the STAR documentation). In additional, household-level weights, raked to the projected population of the STAR districts in 2015, are included. PTRACK in an individual-level file that contains identifiers for each survey round, demographic and interview information, and individual-level weights (also raked to the projected 2015 population by age, sex, district, and urban-rural status). To facilitate using the data, HTRACK and PTRACK for STAR6 carry over a small amount of information from previous rounds of data.  PTRACK contains a measure, agebg_06, that represents our best guess of a person's age at interview (based on comparisons of information across multiple sources). There is also a variable "sex" that represents what we believe a person's biological sex to be. Note that in some cases agebg_06 or sex may differ from the value recorded in the household roster or the book cover.