

*The Fragile Families and Child Wellbeing Study changed its name to The Future of Families and Child Wellbeing Study (FFCWS). Due to the issue date of this document, FFCWS will be referenced by its former name. Any further reference to FFCWS should kindly observe this name change.*

# User's Guide for the Fragile Families and Child Wellbeing Study Public Data, Baseline

November 2018

***Bendheim-Thoman***  
***Center for Research on Child Wellbeing***  
Wallace Hall  
Princeton University  
Princeton, NJ 08544  
<https://crcw.princeton.edu/>

***Columbia Population Research Center***  
1255 Amsterdam Avenue, Room 715  
Columbia University  
New York, NY 10027  
<http://cupop.columbia.edu/>

Prepared by the staff at the Bendheim-Thoman Center for Research on Child Wellbeing (CRCW), Princeton University. For more information about FFCWS, please visit our website at <https://fragilefamilies.princeton.edu/> or email [ffdata@princeton.edu](mailto:ffdata@princeton.edu).

## Table of Contents

0. Study Overview.....	5
0.1 The Core Study .....	5
<i>Table 1: Timeline of the FFCWS Core Study</i> .....	6
0.2 Collaborative Studies.....	6
0.3 National Sample versus Full Sample .....	6
0.4 Data Availability .....	7
0.4.1. Public data .....	7
0.4.2. Contract data.....	7
0.5 Documentation.....	7
1. Baseline Components.....	8
1.1. Funders and Study Administration .....	8
1.2. Surveys and Instruments.....	8
<i>Table 2: Baseline Components and their Sample Sizes</i> .....	8
2. Data Collection Procedures.....	9
2.1. Data collection Procedures - Core Study.....	9
A. Sampling Cities and Choosing Hospitals.....	9
B. Sampling Births.....	10
C. Screening Mothers .....	10
D. Mothers' and Fathers' Eligibility .....	11
E. Interviewing Eligible Mothers.....	11
F. Locating and Interviewing Fathers.....	12
3. File Contents and Structure.....	14
3.1. Variable Structure .....	14
Table 3: Variable name structure (survey variables and weights) .....	14
3.2. Constructed Variables.....	14
3.3. Survey Variables .....	14
<i>Table 4: Survey Instruments in Baseline</i> .....	15
3.4. Key Identifier .....	15
3.5. Variable Label.....	15
3.6. Variable Response .....	15
<i>Table 5: Missing Data Codes</i> .....	16

3.7. Open-Ended Response Codes.....	16
4. Data Cleaning.....	17
5. Weights.....	18
6. Introduction to Topics from the Data.....	19
Table 6: Major topics in Baseline by survey instrument .....	19
7. Paradata.....	20
7.1. Constructed Variables - Age .....	20
7.2. Constructed Variables - Samples Flags.....	20
7.2.1. Interview completion flags .....	20
7.2.2. Status flags .....	20
<i>Table 7: Constructed variables with administrative information.....</i>	<i>21</i>
8. Finances .....	22
Table 8: Subtopics in Finances in Baseline by survey instrument.....	22
8.1. Constructed Variables - Household Income .....	22
8.2. Constructed Variables - Household Income Imputation Flags.....	23
8.3. Constructed Variables - Poverty Measures.....	23
9. Health and Health Behavior .....	24
Table 9: Subtopics in Health and Health Behavior in Baseline by survey instrument.....	24
9.1. Constructed Variables - Low birth weight .....	24
9.2. Scale - Father’s Depression .....	25
9.2.1. Variables.....	25
9.2.2. Modifications .....	25
9.2.3. Scoring.....	25
10. Romantic Relationships.....	26
Table 10: Subtopics in Romantic Relationships in Baseline by survey instrument.....	26
10.1. Constructed Variables - Mother’s relationship with child’s father .....	26
<i>Table 11: Constructed variables about parent’s romantic relationships.....</i>	<i>26</i>
11. Parenting .....	27
Table 12: Subtopics in Parenting in Baseline by survey instrument.....	27
12. Legal System.....	28
Table 13: Subtopics in Legal System in Baseline by survey instrument.....	28
12.1. Constructed Variables - Father in Jail .....	28
13. Housing and Neighborhood.....	29

Table 14: Subtopics in Housing and Neighborhood in Baseline by survey instrument ..	29
Table 15: Constructed variables for household composition .....	29
14. Employment.....	30
Table 16: Subtopics in Employment by survey instrument.....	30
14.1. Open-Ended Response Codes - Occupations .....	30
15. Education .....	32
Table 17: Subtopics in Education by survey instrument .....	32
15.1. Constructed Variables - Parent’s Education .....	32
16. Demographics .....	33
Table 18: Subtopics in Demographics by survey instrument.....	33
16.1. Constructed Variables - Parent’s Race/Ethnicity .....	33
17. Other Topics in Baseline.....	34
Table 19: Other topics and subtopics in Baseline by survey instrument.....	34

## 0. Study Overview

The Fragile Families and Child Wellbeing Study (FFCWS) was initiated to address four questions of great interest to researchers and policy makers:

1. What are the conditions and capabilities of unmarried parents, especially fathers?
2. What is the nature of the relationships between unmarried parents?
3. How do children born into these families fare?
4. How do policies and environmental conditions affect families and children?

The FFCWS follows a cohort of 4,898 children born in the U.S. between 1998 and 2000 and includes an over-sample of non-marital births. The sample includes children born in twenty large, U.S. cities (defined as populations of 200,000 or more). Sixteen of the twenty cities were selected using a stratified random sample of U.S. cities with populations of 200,000 or more grouped according to their policy environments and labor market conditions. These cities comprise the nationally-representative sample. See the sample design paper<sup>1</sup> for details on the selection of cities, hospitals, and births.

### 0.1 The Core Study

The Core Study consists of interviews with both mothers and fathers at the child's birth and again when children are ages one, three, five, and nine. A child interview and in-home observations and assessments are also included at age nine. The Core follow-up at age fifteen includes interviews with the teen and primary caregiver (PCG) as well as in-home observations and assessments.

The parent/PCG interviews collect information on attitudes, relationships, parenting behavior, demographic characteristics, health (mental and physical), economic and employment status, neighborhood characteristics, and program participation. Many measures overlap with those used in other large-scale studies such as the Infant Health and Development Program (IHDP), Early Head Start, the Teenage Parent Demonstration, and the Early Childhood Longitudinal Study—Birth Cohort 2000 (ECLS-B).

See [the FFCWS metadata website](#) to browse or search the full list of FFCWS variables. Table 1 below shows the dates of each wave of data collection.

For the remainder of this Guide, we will refer to the follow-up waves of data collection in reference to the child's age. For example, we will refer to the wave focused upon in this guide as "Baseline" (which is wave 1 in the data file).

---

<sup>1</sup> Reichman et al, "[The Fragile Families and Child Wellbeing Study: Sample and Design](#)" Children and Youth Services Review, 2001, Vol. 23, No. 4/5

*Table 1: Timeline of the FFCWS Core Study*

Wave	Age	Years
1 - Baseline	Birth	1998 – 2000
2	Age 1	1999 – 2001
3	Age 3	2001 – 2003
4	Age 5	2003 – 2006
5	Age 9	2007 – 2010
6	Age 15	2014 – 2017

## 0.2 Collaborative Studies

The Baseline Wave collaborative study, Fragile Families and Child Health, extracted **medical records** that contain information on the mother’s pregnancy and delivery, and the child’s health at birth. This data was collected from approximately 75% of mothers, and is available through the contract data process.

For more details on the collaborative studies at each wave, see that wave’s User Guide or find a [list of all current and completed collaborative studies](#) on our website.

## 0.3 National Sample versus Full Sample

There are 20 cities in the full Fragile Families sample. Sixteen of these cities were selected via a stratified random sample and comprise the “national” sample. For each wave of data and for each unit of analysis (mother, father, couple), users can weight the data up to two different populations – the national level<sup>2</sup> or the city level. Applying the national weights makes the data from the 16 randomly selected cities representative of births occurring in large U.S. cities (the 77 U.S. cities with populations over 200,000 in 1994) between 1998 and 2000. Applying the city-level weights makes the data from all 20 cities in the sample<sup>3</sup> representative of births in their particular city in 1998, 1999, or 2000, depending on the year in which the baseline data collection took place for that city.

The public use data do not contain the geographic identifiers needed to construct the stratum and primary sampling unit (PSU) variables necessary for using a Taylor Series methodology to estimate variances (except through a restricted use contract)<sup>4</sup>. Therefore, the public use data files contain a basic weight and a set of replicate weights. The replicate weights are used in place of the stratum and PSU variables. The replicate weights mask the locations of respondents, while still allowing for estimation of variance. If you are using the public use datasets, you will need to use the replicate weights to get estimates of variance for the sample. Applying the basic weight without

<sup>2</sup> The term national refers to all 77 U.S. cities with 1994 populations of 200,000 or more

<sup>3</sup> There are 109 cases in the data file that were not randomly selected for the core sample (some were randomly selected to be part of a separate study – the TLC3 study) and do not have national sample or city sample weights. Data users can identify and remove these cases using the weights sample flags (cm1citsm = 0 at the Baseline Wave).

<sup>4</sup> Please note that data users who have access to the geographic identifiers may still want to use the replicate weights for their estimates. Using the replicate weights will likely yield similar standard errors (at least for cross-sectional estimates) as the alternative method.

the replicate weights will give you comparable point estimates, but will yield incorrect variance estimates. A brief introduction to the weights available for the public data files is available in the documentation memo "[Fragile Families & Child Wellbeing Study: A Brief Guide to Using the Weights for Waves 1-6](#)." For detailed information on the construction of the weights, see "[Fragile Families & Child Wellbeing Study: Methodology for Constructing Mother, Father, and Couple Weights for Core Telephone Surveys](#)".

## 0.4 Data Availability

There are two types of data available to data users.

### 0.4.1. *Public data*

Currently, Baseline, Year 1, Year 3, Year 5, Year 9 and Year 15 public data are available through the Princeton University [Office of Population Research \(OPR\) data archive](#). To access these data, researchers must complete a brief application and a 25-word abstract about their research project. These files are available in Stata, SPSS, or SAS format and can be downloaded as one combined file (ff\_allwaves\_2018) or in six separate files by wave (ff\_wave1\_2018).

### 0.4.2. *Contract data*

Contract data require a more formal application due to the sensitive nature of the items available. Contract data available includes files, such as a **geographic file** with variables for the focal child's birth city, mother's and father's state of residence at each interview, and stratum and PSU (note: replicate weights are available on the public file in lieu of these), a set of **contextual characteristics** of the census tract at each wave, **medical records data** for mothers and children from the birth hospitalization record, a **school characteristics file** based on National Center for Educational Statistics data, a **labor market and macroeconomic file** with data on local employment and national consumer confidence at each wave, and a **genetic data file** with candidate genes and telomere length.

For further detail regarding the content of the contract data and the application process for its access, please [visit our website](#).

## 0.5 Documentation

The remainder of this guide will provide a detailed overview of the **Baseline Wave of the FFCWS**.

For User Guides for other waves of the FFCWS and further documentation including questionnaires and codebooks for each interview or weights documentation, see the [Documentation page](#) on our website.

## 1. Baseline Components

The Baseline Wave of the Fragile Families and Child Wellbeing Study (FFCWS) contains one main component:

1. The FFCWS Core Study [a.k.a. “Core Study”] (includes mother and father interview)

The Baseline public data file (ff\_baseline\_2018) includes data from the Core Study interviews.

### 1.1. Funders and Study Administration

Funding for the Core Study at Baseline was provided through grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD).<sup>5</sup> A consortium of government agencies, as well as private and non-profit organizations also provided funding support for the Baseline Wave.

The National Opinion Research Center (NORC) was subcontracted to conduct Fragile Families data collection from 1999 through 2000, including collection of baseline data for the first seven cities. In 1999, we contracted Mathematica Policy Research, Inc. (MPR) to complete baseline data collection in the remaining thirteen cities.

The FFCWS Core Study was a joint effort by Princeton University’s Center for Research on Child Wellbeing (CRCW) and Center for Health and Wellbeing (CHW), the Columbia Population Research Center (CPRC) and the National Center for Children and Families (NCCF) at Columbia University.

### 1.2. Surveys and Instruments

This wave involved two surveys or instruments – the core mother and father surveys -- as listed in Table 2. This table also includes the sample sizes for each survey or instrument. For explanations of the variation in sample size, see the sections below on Eligibility and Data Collection Procedures.

*Table 2: Baseline Components and their Sample Sizes*

<b>Study</b>	<b>Surveys and Instruments</b>	<b>N</b>
Core Study	Mother Survey	4,898
	Father Survey	3,830

---

<sup>5</sup> award numbers R01HD36916 (Core)



## 2. Data Collection Procedures

### 2.1. Data collection Procedures - Core Study

The baseline wave of data collection took place from 1998 to 2000. The Fragile Families Study uses a stratified random sample of the 77 U.S. cities having populations of 200,000 or more. Cities were stratified into nine types of environments according to the generosity of welfare benefits, the degree of child support enforcement, and the strength of the local labor market. The study design includes baseline interviews conducted with recent mothers in the maternity wards of the 75 hospitals within 20 U.S. cities included in the study.

At baseline, eligible mothers were asked to identify the father of the child, and fathers were interviewed in person during hospital visits or by telephone. Nearly all of the baseline mother interviews took place in person and over three-quarters of father interviews were in person—the remainder were interviewed over the telephone.

#### *A. Sampling Cities and Choosing Hospitals*

Cities were scored to identify those with extreme values for each of the policy and labor market conditions. One city was randomly selected from each of the eight types of extreme environments (e.g., one city with generous welfare benefits, strict child support enforcement, and a strong labor market, another city with generous welfare benefits, strict child support enforcement, and a weak labor market, and so on). Eight additional cities were randomly selected from the group of cities with moderate policy or labor market conditions. Four additional cities of specific interest to researchers/funders were also included in the study.

In 5 cities, we were able to interview in all birthing hospitals within the city. In 13 cities, with a few exceptions, we rank-ordered the birthing hospitals from those that had the most nonmarital births to those that had the least nonmarital births. In a given city, we chose hospitals in order starting with the largest hospital in terms of the number of nonmarital births until 75 percent of the non-marital births in the city were covered. In two cities, due to their size, we used a simple random sample to select hospitals for the study. See Reichman et al<sup>6</sup> for further detail on the hospital selection process.

Before fielding the survey, we obtained approval to interview recent parents from each sampled hospital. A hospital sponsor (usually a clinician) was recruited to serve as the local Principal Investigator, and to assist in obtaining human subjects approval from the hospital's Institutional Review Board (IRB). With the sponsor's assistance, we submitted a formal request to conduct the study to the IRB. This typically required submission of the survey protocol, participant consent forms, survey instruments, and certificates of human subjects training from each Principal Investigator. Once institutional approval was obtained from each hospital, field staff, trained by the data collection subcontractor, began sampling mothers.

---

<sup>6</sup> Reichman et al, "[The Fragile Families and Child Wellbeing Study: Sample and Design](#)" Children and Youth Services Review, 2001, Vol. 23, No. 4/5

### *B. Sampling Births*

The study was designed to oversample unmarried births, while selecting a smaller sample of married births for comparison. Quotas for the number of unmarried and married parents to be interviewed were set at each hospital, to mimic the hospital's 1996-7 unmarried birth rates. Interviewers attempted to complete interviews with all eligible couples until the quota for married parents was reached. Thereafter, they screened for marital status and only attempted to interview unmarried parents.

The sample frame for each hospital was simply the list of all possible maternity beds. To ensure that each bed had an equal chance of being sampled, maternity rooms were listed in numerical order, with rooms having more than one bed appearing on the list more than once. For example, the list included first the "A" beds in a room (such as beds near the window), then "B" beds. Beds were pre-chosen by their numerical order, regardless of occupancy. If a bed became occupied out of order, it was not selected until it fell back into the sample during the next round of ordered selection. If a bed was empty, the interviewer moved on to the next bed.

For the Baseline survey, Mathematica Policy Research, Inc. (MPR) recruited five to six experienced field interviewers for each city in which the study was conducted. Interviewers were trained in-person on a city-by-city basis. Prior to interviewing in each hospital, MPR site coordinators and field managers ascertained the hospital's visiting hours, the best times to interview, and the locations of private spaces such as hospital waiting areas that could be used for interviewing. Field staff worked with hospital staff to finalize procedures for identifying eligible mothers and obtaining lists of maternity beds. Interviewing for all hospitals was done in accordance with the hospital's specific rules and procedures, as indicated in the hospital fact sheets. A few hospitals requested that the study introduction and request for participation be made by the hospital nursing staff. Informative brochures explaining the purpose of the study were also provided for the mothers' review. Mothers were told that participation in the study was voluntary and, in hospitals where financial incentives were permitted, that they would receive twenty dollars for participating. If a mother agreed to participate, a field interviewer administered the screening instrument to determine the mother's eligibility for the study. All survey materials, including brochures, consent forms, screening instruments and questionnaires, were available in both English and Spanish.

### *C. Screening Mothers*

Prior to administering the Baseline survey, interviewers determined whether or not the mother was eligible to participate by administering a screening instrument that consisted of eight questions. The instrument included questions on whether the mother was married to the father of the baby, if she was 18 years, or older and whether she was planning to place her baby for adoption as well as questions on the status of the father. The screener also collected information on when and if the mother expected the father to visit.

Eligibility requirements were based on the analytical goals and design of the study, including the need to interview both a mother and father of a child who would be

residing with at least one of those parents over the next five years. For example, mothers who were placing their baby for adoption and mothers who reported that the child's father was deceased were considered ineligible. Mothers were also considered ineligible if they were minors in hospitals that did not permit inclusion of minors in the study. Additionally, mothers could be considered ineligible for logistical reasons, including discharge from the hospital before screening and inability to participate in an interview in English or Spanish. Since quotas for number of married and unmarried participants were determined at the start of the study, a married mother screened after the quota for married parents had been reached was also considered ineligible.

Upon completing of the screener and determining the eligibility of a mother, an interviewer reviewed a participation consent form with the parents. Interviewers made sure respondents understood each section of the consent form and gave respondents a chance to ask questions. Respondents were then asked to sign the consent form.

If the mother was considered ineligible to participate in the survey based on the screening instrument, she was informed that an interview would not be needed and was thanked for her time.

In some cases, a mother left the hospital after she had completed a screener but before an interview could be administered. The only circumstance under which a screened and eligible mother could be interviewed after leaving the hospital was if she had already signed a consent form and the father had been interviewed. Under this circumstance the mother was called to complete the interview by telephone.

#### *D. Mothers' and Fathers' Eligibility*

The baseline response rate for mothers measures the percent of all eligible mothers giving birth in the hospital during the data collection period who completed a baseline interview. In order to calculate response rates for married and unmarried mothers in the Fragile Families Study, we grouped the mothers by marital status, screening status, and eligibility status. It was possible for a mother's marital status or eligibility status, or both, to be recorded as unknown in the dispositions from our survey contractors. Marital status and eligibility were imputed (according to the American Association for Public Opinion Research (AAPOR) guidelines) when each characteristic was unknown.

A father was considered eligible to be included in the Fragile Families Study only if the mother of his baby completed a baseline interview (and had, therefore, had been screened and was eligible). See Sample Sizes and Response Rates in Section III for mothers' and fathers' response rates at baseline and each follow-up.

#### *E. Interviewing Eligible Mothers*

Before the baseline mother interview was administered, field staff obtained a signed informed consent form. Interviewers were instructed to allow the mother to read the consent form (or to read it to her if preferred) and to give the mother an opportunity to ask questions about her participation in the study. The mother interview took, on average, 42 minutes to complete, and was attempted immediately after the screener

unless the father was visiting. If the father was present at the hospital immediately after the mother was screened and found eligible, the father interview was attempted first. This was done since his continued availability at the hospital was considered less certain than the mother's availability. If the father was not present at the hospital, an interview with an eligible mother was attempted immediately after she completed the screening instrument. Interviewers took steps to ensure that both interviews were confidential. Mothers and fathers were not interviewed in each other's presence, and interviewers waited until all visitors left the room before conducting an interview.

Once the mother's interview was completed, the mother was thanked for her participation and provided, when permitted by the hospital IRB, with a check for twenty dollars. If the father had not yet been contacted or interviewed, the mother was asked to provide contact information on the father. A second level of consent was also requested from the mother after the interview was completed. This consent gave permission for interviewers to collect basic information from the medical records of both the mother and her child. The use of medical records allowed verification of information the mother provided during the interview and provided basic medical information such as the child's Apgar scores.

#### *F. Locating and Interviewing Fathers*

Before a baseline father interview could be attempted, it was necessary that the baby's mother complete a screening instrument to determine her eligibility, and that she give her signed consent for participation. The baseline father interview was completed in one of four contexts. In order of preference and efficiency, the father interview could take place:

- 1) In the hospital, while the mother was still in the hospital
- 2) From the hospital by telephone (most often by use of a cellular phone) within one week of the baby's birth
- 3) From MPR's telephone center within two to three weeks of the baby's birth
- 4) In-person at the father's home or other location within approximately one month after the baby's birth

Once the mother interview was completed, field staff asked for the mother's assistance in locating the baby's father. For cost reasons, it was preferable to interview the father at the hospital. Mothers were encouraged to provide father's visiting schedules. If a father could not be interviewed while the mother was still in residence, interviewers made every attempt to interview the father within one week of the birth. Interviewers were provided with business cards that could be given to the mother and passed on to the father. These cards contained the interviewer's local cell phone number, as well as a toll free telephone number to MPR's telephone center in Princeton, NJ. Interviewers were also instructed to attempt to call the father at his home to complete the interview by telephone, and to call the mother at home to ask her assistance in gaining the father's participation.

If a father interview could not be completed within a week of the baby's birth, the case was sent to MPR's telephone center where telephone interviewers dedicated to the survey could attempt to reach the father.

If a father interview could not be completed within two weeks of the baby's birth, the case was referred to a field interviewer for additional in-person attempts. This was a particularly effective method for reaching fathers who had wrong or non-working telephone numbers.

Interviewers were trained to deal sensitively with the situation of unwed parents. When attempting to contact fathers outside the hospital, they were required to keep the specific nature of the study confidential, as some respondents might be living with extended family members who had no knowledge of the baby. In such cases, materials sent to the father's address made no reference to "parents." Once the father was contacted, he was offered the option of meeting in a private location outside of his home or of completing the interview by telephone.

Sixty-six percent of completed baseline father interviews were conducted in the hospital, 20 percent of baseline father interviews were conducted by telephone, and the location of father interview was not recorded for 14 percent of the completed interviews. Baseline father interviews took, on average, 43 minutes to administer and, when permitted by hospital regulations, fathers were offered twenty dollars for their participation.

### 3. File Contents and Structure

#### 3.1. Variable Structure

In the Baseline data, each variable name is unique and uses certain characters, as well as a specific order that will help identify to whom and in which survey the question was asked. All variable names from Baseline begin with an alphabetic character. If the variable name begins with the letter “c”, the variable is constructed (see section 4.2 for more on constructed variables). If not, the variable corresponds to a question asked in a Baseline survey and the first character in the variable name indicates to which instrument the variable corresponds. See Table 3 for a full list of Baseline survey instruments and their prefix letters.

In Baseline variable names, what follows the instrument is the number “1” to indicate the wave of data collection. Furthermore, when the variable is directly associated with the questionnaire (is not constructed), the leaf (the end of the variable) will indicate the section letter and the question number to which the variable corresponds. Below is a deconstructed list of the variable names at Baseline:

*Table 3: Variable name structure (survey variables and weights)*

Variable Name			Survey
Prefix	Wave	Leaf	
m	1	[a-j]1-9	Mother Survey
m	1	natwt citywt *	<i>National/City Weights (for mother)</i>
f	1	[a-j]1-9	Father Survey
f	1	natwt citywt*	<i>National/City Weights (for father)</i>
q	1	natwt citywt*	<i>National/City Weights (for couple)</i>

**Note:** an asterisk (\*) is used to indicate the existence of other characters in the variable name. To provide summaries of the variable names, we used asterisk instead of listing each individual case.

#### 3.2. Constructed Variables

A number of variables were constructed and added to the data set by staff. Variables under this group begin with the letter “c”. Some represent data not otherwise available to the public, and some are merely aggregations of existing data that we provide as a “shortcut” for researchers. Researchers may find these variables useful, but are free to construct them in other ways.

When constructing variables such as age, relationship status, and the household roster, the mother's report was generally used. However, there were a few cases in which the father's report was used to fill in missing information or to correct discrepancies in the mother's report.

**Note:** Raw yes/no questions are typically coded as 1=Yes and 2=No. Constructed yes/no variables are typically coded as 1=Yes and 0=No.

#### 3.3. Survey Variables

Survey variables contain responses to questions asked during a survey and their variable names begin with a letter indicating to which survey they correspond. For a list of survey

instruments and their corresponding prefixes at Baseline, please refer to Table 4. The survey instrument is named for the person answering the questions. Following the prefix and wave, survey variables were named as the item in the instrument. For example, variable m1a4 in the data set contains responses provided to item A4 (*Are you currently married to the father of your new baby?*) in the mother core survey questionnaire.

Other variables generated from survey items with no pre-coded response or items with open-ended responses retained most values as recorded during the interview. We edited a few responses provided for some open-ended items in the public-use file to exclude any part of information which could potentially suggest the identification of the respondent.

*Table 4: Survey Instruments in Baseline*

instrument	instrument description
m	Mother Survey
f	Father Survey
q	Couple (used only as weights)

### 3.4. Key Identifier

Family ID (idnum) is the key identifier on the file for merging and sorting. idnum is the random family case ID that links the biological parents of the child at baseline, and in each subsequent wave, links each additional file to the family sampled at baseline. idnum is a string variable consisting of 4 characters. Because the idnum identifier remains fixed throughout the waves, it can be used to merge data from any wave of the study.

### 3.5. Variable Label

Variable labels in the data and codebook correspond as closely as possible to the questions in the questionnaire; however, for formatting reasons some of the questions have been modified or abbreviated in the labels. Please see the questionnaire for official question wording and response categories.

### 3.6. Variable Response

All variables also have value labels describing valid and missing responses. In addition to the listed response categories in the questionnaire, each variable (including continuous variables) can have any of the following nine negative values that indicate missing data:

*Table 5: Missing Data Codes*

<b>Code</b>	<b>Label</b>
-1	Refuse
-2	Don't know
-3	Missing (due to technical error)
-4	Multiple answers
-5	Not asked (not in survey version)
-6	Logical Skip
-7	Not applicable
-8	Out-of-range
-9	Not in wave

### 3.7. Open-Ended Response Codes

Free response questions (open-ended questions) were coded by staff. Codes were assigned by two staff members working independently and these codes were reconciled by a third staff member.

When appropriate, open-ended responses were recoded into the main response categories of the questions. Open-ended responses that did not fit into the existing response categories were recoded into new categories in the 100 range (101, 102, etc.) if there were 10 or more similar responses. Cases that indicate an “other” but were vague or unique remain coded simply as “Other (not specified).”



#### **4. Data Cleaning**

For data derived from the Core Study, limited data cleaning was performed on the files. Some values were recoded to -8 “out of range” and minor changes were made to earnings, income, household roster, ages, etc. if the decision was clear cut. If not, data was left for the user to decide how to code. Known inconsistencies across variables remain in the data for users to consider in their analysis.

## 5. Weights

The Fragile Families sample was selected using a complex sample design, where the sample members were not selected independently and were not selected with equal probabilities. For instance, non-marital births were oversampled. Therefore, Mathematica Policy Research created a set of weights to adjust for the sample design (probability of selection), non-response at baseline, and attrition based on observed characteristics over the waves.

Public users, who do not have access to the stratum and PSU variables, can use a set of replicate weights to properly estimate variance for the sample. Contract data users can employ the replicate weights or Taylor Series method which incorporates strata and PSU.

A brief introduction to the weights available for the public data files is available in the documentation memo "[Fragile Families & Child Wellbeing Study: A Brief Guide to Using the Weights for Waves 1-6](#)." For detailed information on the construction of the sample weights, please read "[Fragile Families & Child Wellbeing Study: Methodology for Constructing Mother, Father, and Couple Weights for Core Telephone Surveys](#)".

## 6. Introduction to Topics from the Data

Baseline data covers a range of topics throughout surveys administered to the focal child's biological mother and biological father. Below are topics covered in Baseline by survey instrument.

Table 6: Major topics in Baseline by survey instrument

Topics	m	f
Attitudes and Expectations	X	X
Childcare	X	
Demographics	X	X
Education and School	X	X
Employment	X	X
Family and Social Ties	X	X
Finances	X	X
Health and Health Behavior	X	X
Housing and Neighborhood	X	X
Legal System	X	X
Paradata and Weights	X	X
Parenting	X	X
Romantic Relationships	X	X

The next sections of this User Guide are organized by these topic categories. Within each section, we will list **constructed variables** (created by staff to add shortcuts for data users), followed by **scales** and **concepts** that relate to each topic. We define a scale as a composite measure that is composed of variables within the same construct. By constructing a scale, researchers can indicate the degree or intensity to which respondents adhere to the given construct. Scales are typically derived from an established source or existing study. Information on scoring a scale can be found within each section. Concepts are also aggregations of similar variables; however, we do not provide information on scoring, nor do we treat concepts as validated scales.

Researchers are also encouraged to interrogate the data further and to refer to the questionnaires provided in the [Documentation](#) for more information on the survey content.

## 7. Paradata

The Baseline survey includes variables with information about the interview, also known as paradata. Within the available Baseline paradata is the date (month and year) the interview was administered, the language it was administered in (English or Spanish), and the way in which it was delivered to the respondent (in person or by phone).

**Sample flag variables** were also constructed by staff to help users sort the data by (1) respondent participation in a given survey and, if applicable, their reason for non-response (for fathers only at Baseline), or (2) whether the respondent belongs to the nationally-representative or city-representative sample. The rest of this chapter will highlight specific constructed paradata variables which are provided in the Baseline data. For a full list of constructed variables see Table 7.

### 7.1. Constructed Variables - Age

Age is recorded from the Core Baseline survey for mother, father and child and can be retrieved through the constructed variables: **cm1age** (mother's age at the interview), **cf1age** (father's age at the interview), **cm1b\_age** and **cf1b\_age** for the child's age at the mother and father interview, respectively. Data users should note that the child constructed age in years variable was rounded up or down to the nearest year, based on the calculated age in months.

### 7.2. Constructed Variables - Samples Flags

There are two types of sample flags – **interview flags** and **status flags**. Interview flags denote whether a person was interviewed in a particular wave; these include **cm1fint** and **cf1fint**. Status flags provide other important information describing if a case is in a particular subsample. The following lists the sample flags from Baseline (**cm1natsm**, **cf1natsm**, **cq1natsm**, **cm1natsmx**, **cf1natsmx**, **cm1citsm**, **cf1citsm**, **cq1citsm**).

#### 7.2.1. *Interview completion flags*

- **cm1fint** indicates whether father was interviewed, using mother's record(s).
- **cf1fint** indicates whether father was interviewed, using father's record(s).

Cases in which the respondent was not interviewed in the current wave are included in the files but are coded "Not in wave" (-9) for all variables. Therefore, you will need to use these interview flags to subset out appropriate samples.

#### 7.2.2. *Status flags*

- **c\*1natsm** and **c\*1citysm** indicate whether the respondent is in the national sample and/or the 20-cities sample and was interviewed in that wave

Note: There are a small number of cases that do not have weights but have valid survey data and there are a small number of cases that have positive weights, but no survey data because the parent/child was deceased or the child was adopted.<sup>7</sup>

---

<sup>7</sup> for more information see: ["Using the Fragile Families Weights"](#)

*Table 7: Constructed variables with administrative information:*

Constructed Variable	Description of Constructed Variable
c[m f]1age	Mother's/Father's age (years)
c[m f]1b_age	Child's age at time of Mother/Father interview (months)
c[m f q]1citsm	Baseline city sample flag
c[m f q]1natsm	Baseline national sample flag
c[m f q]1natsmx	Baseline national sample flag (excluding one city)
c[m f]1fint	Was father interviewed at Baseline?
c[m f]1intmon	Mother/Father interview month
c[m f]1intyr	Mother/Father interview year
c[m f]1span	Interview conducted in Spanish
c[f]1tele	Interview conducted by telephone
c[m f]1twoc	Two cities flag
c[m f]1lenhr	Total length of interview (hours)
c[m f]1lenmin	Total length of interview (minutes)
cm1med	Mother's hospital records available

## 8. Finances

Questions were asked to the child’s mother and father on household finances at Baseline. This section describes variables related to finances that were constructed by CRCW researchers. Other subtopics in finance were covered across the surveys and where which questionnaire you might find them in is listed in the table below:

**Table 8: Subtopics in Finances in Baseline by survey instrument**

<b>Subtopics</b>	<b>m</b>	<b>f</b>
Child support	X	X
Earnings	X	X
Expenses		X
Financial assets	X	X
Household income/poverty	X	X
Private transfers	X	X
Public transfers and social services	X	X

### 8.1. Constructed Variables - Household Income

Household income measures were constructed for mothers and fathers, but users should review the following information regarding the imputation and construction process carefully before deciding how and whether to use these variables.

- **cm1hhinc** and **cf1hhinc** are mother and father’s household income at Baseline, respectively.

We provide constructed household income measures but users should carefully consider how/whether to use these variables. Please review the following information carefully.

Baseline household income (cm1hhinc, cf1hhinc - total income earned before taxes) was collected in categorical form. About 25 percent of respondents were missing data. While we provide an imputed Baseline income variable, data users should be aware of the level of missing data and the method of imputation of these data. For those who provided bracketed household income at baseline, we imputed the mean value of the bracket. The “mean” of the top bracket was calculated as the mean CPS value by city, marital status, and year of interview. For married and cohabiting couples, we used mother reports of income if available; otherwise, we used father report if mother report was missing. If neither parent reported income, household income was imputed using Stata’s regression-based impute command and included the following covariates for mothers and fathers: city, age, years of education, race/ethnicity, earnings, immigrant status, employed last year, hours worked, total adults in household, earnings, received welfare, and marital status. For couples who were not married or cohabiting, we used the mother/father report if available; otherwise, missing data was imputed using the same method and covariates (with the exception of marital status) as was used for married and cohabiting couples. For father constructed baseline

household income, mother reports were used if the couple was married or cohabiting. Users can consider their own imputations for alternate constructions.

## 8.2. Constructed Variables - Household Income Imputation Flags

- **cm1hhimp** and **cf1hhimp** indicate which parent reported income and which parents have imputed income.

*Please note* that if parents reported a range of income in brackets, they are not flagged as having imputed data. Users can examine the raw variables to determine who had detailed/bracketed data. Because those reporting bracketed data are assigned the mean of the bracket and those reporting more missing data were imputed (unconstrained) there is more variance in the imputed data than in the reported data. Users can consider alternate imputation strategies.

## 8.3. Constructed Variables - Poverty Measures

- **cm1inpov** and **cf1inpov** indicate the poverty ratio. The poverty ratio is the ratio of total household income, as defined in **c\*1hhinc**, to the official poverty thresholds, designated by the U.S. Census Bureau.
- **cm1povca** and **cf1povca** indicate the poverty categories by transforming the poverty ratios into categorical variables.

The thresholds in **c\*1povca** vary by family composition and year. At each wave, we used the poverty thresholds for the year preceding the interview. We calculated separate thresholds based on mother and father reports of household size and composition. However, calculations for married/cohabiting mothers and fathers rely on mother reports of household size and composition. A small number of missing values (don't know, refused) were treated as 0 in household membership counts.

The imputation flags created for the household income variables also refer to the poverty variables.

Please visit <https://www.census.gov/topics/income-poverty/poverty/guidance/poverty-measures.html> for detailed information about poverty thresholds.

## 9. Health and Health Behavior

The Baseline Wave surveys contain questions on focal child’s birth and health, as well as mothers’ and fathers’ overall health, substance use, and health-related work limitations.

Table 9: Subtopics in Health and Health Behavior in Baseline by survey instrument

<b>Subtopics</b>	<b>m</b>	<b>f</b>
Fertility history	X	X
Health behavior		X
Health care access and insurance	X	X
Height and weight	X	
Mental health	X	X
Physical health	X	X
Substance use and abuse	X	X

### 9.1. Constructed Variables - Low birth weight

The **cm1lbw** variable identifies babies weighing less than 2,500 grams at birth as low-birth-weight. The variable (cm1lbw) is coded one (1) for any baby weighing less than 2,500 grams at birth, and zero (0) for babies who were not low-birth-weight. Note: Gestational age was not available. Since the cut-off point for low birth weight is 5 lbs and 8 ounces, cases that reported the baby’s weight to be 5 pounds, but were missing the ounces in, are coded as -3 (“missing”) on cm1lbw. In the case of multiple births, the variable cm1lbw is coded as -6 (“skipped”) because there are different standards for determining low birth weight(s) in multiple births. Users who want a more precise measure of low birth weight (and other birth health measures) can consider applying for the medical records data via a restricted use contract.



## 9.2. Scale - Father's Depression

### 9.2.1. *Variables*

*Father questions: f1g9a-f1g9l* (12 variables)

These items are drawn from the Center for Epidemiologic Studies Depression Scale (CES-D).<sup>8</sup>

### 9.2.2. *Modifications*

12 of the original 20 items were included in the Baseline Father Survey. The original CES-D items include values on a five-point scale ranging from “not at all or less than 1 day last week” to “nearly every day for two weeks.” In the version included in the Baseline Father Survey, items refer to how many days in the past week they felt this way.

### 9.2.3. *Scoring*

There is no established scoring method given that a subset of the questions were administered.

---

<sup>8</sup> Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement*. 1977;1:385-401.

## 10. Romantic Relationships

A number of questions were asked during the Baseline mother and father surveys to understand the parent's romantic relationship with one another. Questions were asked regarding their relationship quality with their partner (i.e. communication, supportiveness, cooperation, intimate partner violence), and their relationship status (whether they are married, cohabiting, dating, no longer together). Constructed variables regarding their relationship status was made by staff.

Table 10: Subtopics in Romantic Relationships in Baseline by survey instrument

<b>Subtopics</b>	<b>m</b>	<b>f</b>
Relationship Quality	X	X
Relationship Status	X	X

### 10.1. Constructed Variables - Mother's relationship with child's father

- **cm1relf** mother's reported romantic relationship with child's father at Baseline

Mother's reported romantic relationship with baby's father at the child's birth (cm1relf). The relationship status classification is based on information reported by mothers about their marital status (m1b2), cohabitation status (m1b8, m1b20 and m1e1) and how they describe their current relationship with the baby's father (m1b3). Mothers are considered married for cm1relf if m1b2=1. For unmarried mothers (defined as m1b2=2 or m1b2=missing because father is "unknown"), m1b3 and m1b8 are cross-tabulated: those cohabiting and "steady" or "on & off" are classified as cohabiting on cm1relf; those not cohabiting are classified as visiting (romantic, non-cohabiting). If m1b20 and m1e1 (household roster) are used sequentially to determine whether she is cohabiting. Eight cases that are missing information on m1b3 due to a refusal or any other reason are coded as missing (-3) on cm1relf. Three cases in which the mother reported "father unknown" but a complete father interview is available are coded on cm1relf according to father report.

*Table 11: Constructed variables about parent's romantic relationships*

<b>Constructed Variable</b>	<b>Description of Constructed Variable</b>
cm1cohf	Mother living with (not married) child's father at Baseline
cf1cohm	Father living with (not married) child's mother at Baseline
cm1marf	Mother married to baby's father at Baseline
cm1relf	Mother relationship with father at Baseline

## 11. Parenting

Questions were asked to the mother and father at Baseline regarding the parents' expectations about child-rearing. Details such as whether the child will have the father's last name, if his name will be on the birth certificate, and who the child looks like were also collected.

Table 12: Subtopics in Parenting in Baseline by survey instrument

<b>Subtopics</b>	<b>m</b>	<b>f</b>
Parent-Child Contact		X
Parenting Abilities	X	X
Parenting Behavior		X

## 12. Legal System

At Baseline, respondents were asked whether the other parent was incarcerated or had been incarcerated and whether paternity for the child had been established.

Table 13: Subtopics in Legal System in Baseline by survey instrument

<u>Subtopics</u>	<u>m</u>	<u>f</u>
Criminal Justice Involvement	X	X
Paternity	X	X

### 12.1. Constructed Variables - Father in Jail

**cm1finjail, cf1finjail, cm1ffinjail**

The constructed jail variables for mother report of father in jail, father report of his own jail, and mother/father combined reports of father in jail. The jail variables maximize reports of fathers' jail status based on information in the core files and from disposition reports. The variables are coded as 0 for not in jail/never in jail and 1 for in jail/ever in jail. We did not code cases "not in wave" on these variables; instead, missing values represent no information available on jail status.

### 13. Housing and Neighborhood

At Baseline, mothers and fathers were asked for a housing roster which was used to plot the number of people in the home, what relationship the respondent had to each person, how old each person is and whether they were working. In addition, respondents were asked how long they lived in their neighborhood, whether they felt safe in their neighborhood, and about their living arrangements and receipt of housing assistance.

**Table 14: Subtopics in Housing and Neighborhood in Baseline by survey instrument**

<b>Subtopics</b>	<b>m</b>	<b>f</b>
Child Living Arrangements	X	
Home Environment	X	X
Household Composition	X	X
Housing Status	X	X
Parents' Living Arrangements	X	X
Residential Mobility	X	X
Neighborhood Conditions	X	X

*Table 15: Constructed variables for household composition*

<b>Constructed Variable</b>	<b>Description of Constructed Variable</b>
c[m f]1adult	Number of adults 18 or over in household
c[m f]1kids	Number of children under 18 in household
cm1cohf	Mother living with (not married) child's father at Baseline
cf1cohm	Father living with (not married) child's mother at Baseline
c[m f]1cohp	Mother/father living with (not married) new partner at Baseline
c[m f]1gdad	Grandfather present in household
c[m f]1gmom	Grandmother present in household

## 14. Employment

At Baseline, questions were asked regarding the child’s mother’s and father’s employment. In the subtopic traditional employment, fathers were asked about their job type, their work schedule and the employment status of other members of their household. Mothers were asked about their past employment, the father’s employment, and the employment status of other members of their household. In the non-traditional work subtopic, fathers were asked about their non-traditional job (including working for self, “hustles”, and other work) work schedule in the last year (both frequency and type of work) and mothers were asked if the father did any non-traditional work. In the unemployment subtopic, fathers were asked about their current employment status, whether and how long they’ve been looking for a regular job, if they received income from unemployment or disability, as well as whether they had never regularly worked. Mothers were also asked if they had never regularly worked, if they received income from unemployment or disability, and whether the father never regularly worked.

Table 16: Subtopics in Employment by survey instrument

<u>Subtopics</u>	<u>m</u>	<u>f</u>
Traditional work	X	X
Non-traditional work	X	X
Unemployment	X	X

### 14.1. Open-Ended Response Codes - Occupations

For **traditional employment**, we constructed an occupation variable for fathers (f1j7bc) based on the 3 digits codes from the U.S. Bureau of Labor Statistics (BLS) Occupational Classification System by Major Occupational Groups. These categories are summarized below:

- 101 – Professional, Technical, and Related Occupations
- 102 – Executive, Administrative, and Managerial Occupations
- 103 – Sales Occupations
- 104 – Administrative Support Occupations, including Clerical
- 105 – Precision Production, Craft, and Repair Occupations
- 106 – Machine Operators, Assemblers, and Inspectors
- 107 – Transportation and Material Moving Occupations
- 108 – Handlers, Equipment Cleaners, Helpers, and Laborers
- 109 – Service Occupations, except Private Household
- 110 – Unspecified
- 112 – Military
- 113 – Farming/Agriculture
- 114 – Self-employed
- 115 – Various Jobs

For **non-traditional employment** (e.g. work in own business (f1j13b) and other source of income (f1j13d)), variables were coded using a slightly different set of categories following the classifications described by *Occupational Classification System by Major Occupational Groups* (though these code numbers differ slightly).

101 – Artists and Athletes: includes athletes, photographers, artists, musicians. This category is based on a Board of Labor Statistics sub-grouping.

102 – Administrative Support: to include clerical jobs, bookkeepers, and people working for temp agencies.

103 -- Sales

104 – Construction and Precision Trades: includes jobs related to building and home improvement (brickmasons, carpet installers, drywallers, painters, carpenters, etc) as well as the respondent who said he makes uniforms. This is based on the BLS Major Occupational Group E with mechanics and repairers removed. (See code 110)

105 – Military

106 – Entertainment: includes escort service, adult entertainment, party services, DJs, and gambling.

107 – Transportation and Delivery

108 – Service Occupations: includes food (restaurants, catering, bartending), health (aromatherapists, personal trainers), and personal services (babysitting, in home care of the elderly, cosmetology). This is based on BLS Major Occupational Group K.

109 – Illegal Activity

110 – Mechanics and Repairers: includes work related to car repair or audio installation. This is the other portion of BLS Major Occupational Group E (most are in code 104).

111 – Real Estate and Finance

112 – Landscaping and Agriculture: includes landscaping, cutting grass, ranching, farming, raising cattle.

113 – Professional: includes educators, lawyers, accountants, architects, information technology jobs, and other professionals. This is essentially BLS Major Occupational Group A without artists & athletes (code 101)

114 – Other: includes responses we could not code into above.

## 15. Education

At Baseline, mothers and fathers were asked about their own highest level of educational attainment as well as other training and schooling they have received.

Table 17: Subtopics in Education by survey instrument

<u>Subtopics</u>	<u>m</u>	<u>f</u>
Educational Attainment/Achievement	X	X

### 15.1. Constructed Variables - Parent's Education

- **cm1edu, cf1edu** mothers' and fathers' education at baseline

Mothers' and fathers' education at Baseline are based on their own reports, but mothers' reports of fathers' education are used for fathers who were not interviewed at baseline or did not report their own education.



## 16. Demographics

At Baseline, mothers and fathers were asked a number of demographic questions including their own age and the ages of others in their household. They were also asked about whether they were both in the United States, if not, where, and what year they came to live in the U.S. (if not born in the U.S.). There are variables for the language the survey was conducted in, the race/ethnicity of each parent, the sex of the focal child, and the gender of each person in the household.

Table 18: Subtopics in Demographics by survey instrument

<b>Subtopics</b>	<b>m</b>	<b>f</b>
Age	X	X
Citizenship and Nativity	X	X
Language	X	X
Race/Ethnicity	X	X
Sex/Gender	X	X

### 16.1. Constructed Variables - Parent's Race/Ethnicity

- **cm1ethrace, cf1ethrace** mothers' and fathers' race/ethnicity

Mothers' and fathers' race/ethnicity are based their own reports, but mothers' reports of fathers' race are used for fathers who were not interviewed at any wave.

## 17. Other Topics in Baseline

The following table includes subtopics within topics that are not explicitly written about in this user guide. For more on these topics, please refer to the survey instruments/questionnaires.

**Table 19: Other topics and subtopics in Baseline by survey instrument**

<b>Topics and Subtopics</b>	<b>m</b>	<b>f</b>
<i>Attitudes and Expectations</i>		
Attitudes/Expectations/Happiness	X	X
<i>Childcare</i>		
Childcare Services and Availability	X	
<i>Family and Social Ties</i>		
Grandparents	X	X
Parent's Family background	X	X
Religion	X	X
Social Support	X	X