Introduction to the Fragile Families Public Use Data

Baseline, One-Year, Three-Year, and Five-Year Core Telephone Data

August 2008

Bendheim-Thoman
Center for Research on Child Wellbeing   ***Columbia Population Research Center***
Wallace Hall   1255 Amsterdam Avenue, Room 715
Princeton University   Columbia University
Princeton, NJ 08544   New York, NY 10027
http://crcw.princeton.edu   http://cupop.columbia.edu/

I.   STUDY OVERVIEW

The Fragile Families and Child Wellbeing Study 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 Study follows a cohort of nearly 5,000 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 20 large, U.S. cities (defined as populations of 200,000 or more). Sixteen of the 20 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 (Reichman et al, "The Fragile Families and Child Wellbeing Study: Sample and Design" *Children and Youth Services Review*, 2001, Vol. 23, No. 4/5) for details on the selection the cities, hospitals, and births. The documentation memo "Fragile Families & Child Wellbeing Study: A Brief Guide to Using the Mother, Father, and Couple Replicate Weights for Core Telephone Surveys Waves 1-4" is also useful for understanding the samples and populations.

A. 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, and five. A nine-year follow-up is in the field from 2007-2009 with a data release anticipated in 2010/2011.[1] The parent interviews collect information on attitudes, relationships, parenting behavior, demographic characteristics, health (mental and physical), economic and employment status, neighborhood characteristics, and program participation. See the questionnaire map (available on our website) for a brief overview of the topics covered in the core telephone interviews. 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 Scales Documentation available on the Fragile Families website for more details on established scales that were used/adapted in the core study.

The baseline interviews were conducted between February 1998 and September 2000. The one-year follow-up interviews were conducted between June 1999 and March 2002.   The three-year follow-up interviews were conducted between April 2001 and December 2003. The five-year follow-up interviews were conducted between July 2003 and February 2006.

B. Collaborative Studies
There are four collaborative studies that will yield public use data files for subsets of the core sample. Each of these studies will release their own data files and documentation.

---

[1] The nine-year follow-up merged the core telephone survey, the in-home study, and a teacher study into one large project. Saliva samples are also being collected in order to study genes associated with learning, behavior, health, and child development.

The first collaborative study is the In-Home Longitudinal Study of Pre-School Aged Children, which includes a primary caregiver survey and in-home assessments. At ages three and five, the child's primary caregiver (typically the child's mother, unless the child lives with the father or a non-parental caretaker) participates in an additional in-depth interview of about an hour that focuses on parenting, child health, and development. This interview, usually conducted in the child's home, is accompanied by a set of direct assessments of parenting, child health, and development.

The second collaborative study is the Child Care and Parental Employment Study, which contributed questions on child care/early education and maternal employment to the core and primary caretaker surveys (so that data is folded into the core and in-home data files) and also conducts child care provider/teacher surveys and direct assessments of child care quality at the three-year follow-up and kindergarten surveys at the five-year follow-up.

The third collaborative study is Fragile Families and Child Health, which extracted medical records that contain information on the mother's pregnancy and delivery and the child's health at birth.

The final collaborative study is the Time, Love, and Cash among Couples with Children (TLC3) study. In this study, interviewers conduct in-depth qualitative interviews over the course of five years with a subsample of 75 romantically involved couples in the Fragile Families survey in three cities - Milwaukee, Chicago, and New York. The interviews began two to three months after the baby's birth, with follow-up waves when the child was approximately 12 months, 24 months and 48-50 months of age. The in-depth, semi-structured interviews cover topics that include parents' relationship with their partners (child's father or a new partner), division of household labor, and ideals and norms about marriage and fatherhood.

II.   DATA AVAILABILITY

There are two types of data available to data users. Public data is available by completing a brief application and a 25-word abstract about your research project. Contract data requires a more formal application due to the sensitive nature of the items available.

A. Public Data
Study questionnaires, documentation, data alerts, responses to frequently asked questions, and a timeline for all public use file expected release dates are available on the Fragile Families web site. Data are available for download from the Princeton University Office of Population Research (OPR) data archive. Currently, baseline, one, three-, and five-year core telephone data are available to the public as well as the three-year in-home data. The data files are packaged in WinZip archives containing SAS, SPSS, and Stata data sets. Please visit our Frequently Asked Questions page for help with downloading, unzipping, and using the data sets.

B. Contract Data
In order to protect the confidentiality of survey respondents, geographic (e.g., city and state) identifiers are not released on the public use data files. This includes the stratum and PSU

variables necessary to estimate variance of weighted estimates using Taylor Series methodology. Replicate weights are provided to public users as an alternative to using Taylor Series. See Section VI. Sample Weights.

Users can apply for access to the following items via a restricted use data contract: geographic identifiers (city/state), stratum and PSU, contextual data at the neighborhood level (see website for a list of variables that are currently available), and medical records data from the child's birth.

See http://www.fragilefamilies.princeton.edu/restricted.asp for more information on the contract data process and requirements.

III.   DOCUMENTATION

A. Using the Questionnaires
There are three types of response sets used in the survey:
1. The choices were read to the respondent (for this type of question possible response categories are in lower case.)
2. Choices were presented to the respondent on a card (indicated by interviewer instructions.)
3. Answers were coded by the interviewer into categories to best correspond to the answer of the respondent (possible response categories are in CAPS.)

Questions in BOLD in the survey are interviewer check questions that summarized information previously gathered to facilitate skip pattern; they are not asked of the respondents. The corresponding variables in the data files reflect this beginning with "INT CHK."

Respondents that replied "don't know" or "refused," or were missing a response in a question that involved a skip pattern were skipped from the subsequent question(s). We annotate the questionnaires with information about known skip problems.

Questionnaire maps for the core and in-home surveys are available on the Fragile Families website.

B. Changes to the Questionnaires
Questions that were added to the survey during fielding
Some questions were not included on all versions of the surveys because they were added or dropped during fielding. Respondents may not have been asked a question because it was not on the instrument at the time they were interviewed. If so, we
- Denote the individual as "NOT ASKED" on a question, which is coded "-5" in the data. (If we could construct the information from other parts of the survey, we did so.)
- Indicate in the instrument that the question is available in "18 Cities Only".
- Retain similar information if available. If a similar, but not parallel question existed in the previous version, we retained those data so that data users could decide how/whether to combine questions. These variables are renamed to include an "X" in their prefix (i.e. **MX1J2** and **FX1K2**). For instance, at baseline, in the first version of the survey we

asked respondents their total income for the previous 12 months, whereas in the second version we asked the amount of income by different sources. We retained the total income amount variables for those who responded to the first version of the survey -- **J1** in the mothers' questionnaire and **K2** in fathers' questionnaire. These variables are renamed to include an X in their prefix (i.e. **MX1J2** and **FX1K2**).

In most instances (but not all), changes to the questionnaire occurred between fielding the first two cities and the subsequent 18. Therefore, we have constructed flag variables (e.g., cm1twoc) to indicate cases are in the first two cities. We also included the letters "TCO" in the variable labels for the "X" variables to indicate that questions were asked in "Two Cities Only". These variables are only included if there are a significant number of responses or if a skip pattern change results. In the questionnaires and data files, these variables are located immediately after the questions measuring the same concept.

Questions that were changed during fielding
Questions that were modified between versions the questionnaires are annotated in the survey instruments. For example, new response categories were added to questions asking about the education of the respondent's biological mother or father at the one-year follow-up (**M2G2, M2G3, F2G2 and F2G3**). These categories ("graduate school" (11) and "some college" (12)) were added to the end of the list of possible response categories. **Please note:** Response categories for these questions are out of the logical order.

If response categories changed in a substantive manner, we recoded respondents' answers into 200 range codes (e.g., 201, 202, etc.) allowing the data user to decide if/how to use these data. This is most relevant for the two cities sample, but can also apply to respondents in other cities. For example, m2b18a has 200 codes for some respondents in other cities as well.

IV.   USING THE DATA

A. Sample Sizes and Response Rates
Below is a table of sample sizes and response rates for each of the three waves. The data files include records for 4,898 families, approximately 3,700 of whom were unmarried at the child's birth. At the one-year, three-, and five-year follow-ups, we attempted to re-interview all mothers interviewed at the child's birth and all fathers of children whose mother we interviewed at the child's birth, even fathers who we missed at baseline (and subsequent waves).[2] See Section VII. Data Collection Protocols for interviewing and locating protocols.

The sample sizes, completion rates, and response rates for each wave are in the table below.

---

[2] We obtain demographic information for fathers who are first interviewed after baseline in section G at the one-year and section H at the three- and five-year follow-ups. Mothers also report on fathers' characteristics at baseline (race, age, education, employment) and subsequent waves (employment, number of children).

	Mother - 20 cities			Mother - national sample			Father - 20 cities			Father - national sample		
	Total	Married at birth	Unmarried at birth	Total	Married at birth	Unmarried at birth	Total	Married at birth	Unmarried at birth	Total	Married at birth	Unmarried at birth
Sample sizes												
Baseline	4789	1141	3648	3442	820	2622	3742	1015	2727	2726	732	1994
One-year	4270	1029	3241	3082	745	2337	3306	932	2374	2409	679	1730
Three-year	4140	1012	3128	2973	734	2239	3225	928	2297	2315	677	1638
Three-year in-home (any)[1]	3288	802	2486	2338	580	1758	n/a	n/a	n/a	n/a	n/a	n/a
Three-year in-home (assessments)[2]	2581	578	2003	1819	407	1412	n/a	n/a	n/a	n/a	n/a	n/a
Five-year	4055	975	3080	2927	717	2210	3087	870	2217	2235	637	1598
Completion rates = complete interview/mother baseline complete in city or national sample												
Baseline	100	100	100	100	100	100	78	89	75	79	89	76
One-year	89	90	89	90	91	89	69	82	65	70	83	66
Three-year	86	89	86	86	90	85	67	81	63	67	83	62
Three-year in-home (any)[3]	79	79	79	79	79	79	n/a	n/a	n/a	n/a	n/a	n/a
Three-year in-home (assessments)[3]	62	57	64	61	55	63	n/a	n/a	n/a	n/a	n/a	n/a
Five-year	85	85	84	85	87	84	64	76	61	65	78	61
Core response rates = (interviewed + obtained all relevant info)/(mother baseline in city or national sample - ineligible)												
Baseline[4]	86	82	87	86	82	87	78	89	75	79	89	77
One-year[5]	90	91	90	91	91	91	74	82	71	74	83	71
Three-year[5]	88	89	88	88	90	88	72	82	69	72	83	68
Five-year[5]	87	86	87	87	88	87	70	78	67	70	79	67

Notes:

The sample sizes are the maximum number of cases you could expect for your analyses before losing cases for item non-response. There are 109 cases in the data file that are not included in this table. Those cases were not randomly selected for the core sample (some were randomly selected only to be part of a separate study – the TLC3 study). These cases do not have national sample or city sample weights. Data users can identify and remove these cases using the weights sample flags (cm1citsm = 0 or incitysm=0).

1 Includes mothers with any available in-home data (telephone, assessment, and/or observation). See in-home documentation for more detail.

2 Includes mothers with any in-home assessments (e.g. child achievement tests). May not have all in-home measures. See in-home documentation for more detail.

3 As percentage of core completes for that wave. Note: not all in-home respondents are mothers.

4 Baseline mother response rates are percentage of eligible mothers approached in the hospitals. Father response rates are relative to completed mother baseline interviews.

5 Core follow-up mother and father response rates are the percentage of completed interviews over the number of mother baseline interviews minus ineligibles at the follow-up. For the purposes of the response rates, deaths and cases not interviewed because child was living outside of the home are treated as completed interviews. Information re: eligibility and nonresponse can be found in c*samp variables and more information about eligibility is documented in Section VII of this guide.

B. Key Identifiers

The identifiers on the file for merging and sorting include a family ID, and mother and father IDs. **idnum** is the random family case ID that links the biological parents of the child at baseline. This is a string variable consisting of 4 characters. **idnum** should be used for merging mothers and fathers data. **mothid1** is the mother's case ID at baseline, consisting of the 4 characters in the **idnum** with an additional "0" at the end. **fathid1** is the father's case ID at baseline, consisting of the 4 characters in the **idnum** with an additional "1" at the end. The **idnum** identifier will remain fixed throughout the waves. The one-, three-, and five-year follow-ups contain copies of the individual identifiers (mothid2, mothid3, mothid4, fathid2, fathid3, and fathid4) to facilitate checking merges.

C. Sample 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. Status flags provide other important information about a case at a particular period (non-response reason, in a particular subsample, etc). The table lists the key sample flags and brief descriptions of these flags follow.

	Baseline	**One**	**Three**	**Five**
	Mother			
Sample flag		cm2samp	cm3samp	cm4samp
National sample flag	cm1natsm	cm2natsm	cm3natsm	cm4natsm
National sample minus one city	cm1nasmx	cm2natsmx	cm3natsmx	cm4natsmx
City sample flag	cm1citsm	cm2citsm	cm3citsm	cm4citsm
Mother interviewed at wave		cm2mint	cm3mint	cm4mint
Father interviewed at wave				
In-Home sample			cm3inhom	cm4inhom
Child Care sample			cm3inccprov	
Different father	cm1fdiff	cm2fdiff	cm3fdiff	cm4fdiff
	Father			
Sample flag	cf1samp	cf2samp	cf3samp	cf4samp
National sample flag	cf1natsm	cf2natsm	cf3natsm	cf4natsm
National sample minus one city	cf1natsmx	cf2natsmx	cf3natsmx	cf4natsmx
Mother interviewed at wave	cf1mint	cf2mint	cf3mint	cf4mint
Father interviewed at wave	cf1fint	cf2fint	cf3fint	cf4fint
	Couple			
National sample flag		cc2natsm	cc3natsm	cc4natsm
City sample flag		cc2citsm	cc3citsm	cc4citsm

Interview flags on the mothers' record(s) indicate whether she was interviewed (**cm*mint**) and whether the father was interviewed (**cm*fint**). Father records also have interview flags for whether he was interviewed (**cf*fint**) and whether the mother was interviewed (**cf*mint**). Cases that were not interviewed in the current wave are included on 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.

Status flags cm*samp and cf*samp provide disposition information about the status of the respondent at those waves. They provide information on eligibility and reasons for non-response such as mother/father/child death between waves. Note: these flags are not available on the baseline files.

There are also flags at each wave that indicate whether the respondent is in the national sample and/or the 20-cities sample and was interviewed in that wave (**c**natsm/c**citysm**). There are also flags on the baseline file that indicate whether the respondent was part of the national/city sample regardless of whether they were interviewed at any given wave (**innatsm/incitysm**). See the next section for more information on these samples.

Note: There are a small number of cases that do not have weights but have valid survey data (see endnote V in "Using the Fragile Families Weights" for more detail) 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 (see Appendix B of "Using the Fragile Families Weights" for more information).

cm3inhom indicates whether the child completed any part of the collaborative study, the In-Home Longitudinal Study of Pre-School Aged Children, at the three-year follow-up.

cm3inccprov indicates whether the family participated in the child care provider collaborative study.

A handful of mothers provided conflicting information over the waves about who is the biological father of the child. The **cm*fdiff** variables flag such cases where the mother indicated the biological father was a different man than had been indicated at earlier waves and for whom we had no reason to doubt this information. However, we cannot determine the accuracy of these reports.

D. 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[3] 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[4] 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.

---

[3] In this memo, the term *national* refers to all 77 U.S. cities with 1994 populations of 200,000 or more.
[4] 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 or **incitysm** = 0).

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)[5]. 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 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 Mother, Father, and Couple Replicate Weights for Core Telephone Surveys Waves 1-4." 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 Waves 1-4".

E. Multiple Births
Variable **cm1numb** at baseline indicates whether or not the focal child was part of a multiple birth. If the mother had more than one baby at the current birth, the variable cm1numb is an indicator of the number of babies born.

F. Notes on Father/Couple Data
At baseline, we attempted to interview both parents as soon after the baby's birth as possible. Most of the data, for mothers as well as fathers, was collected within the first three days after the child's birth. However, we continued to pursue cases we were unable to interview right away. Mothers were interviewed between 0 and 112 days after their baby's birth, with 99 percent occurring within the first week after birth. Fathers were interviewed between 0 and 381 days after their baby's birth, with 77 percent occurring within the first week after birth. Note: Data users are not provided with day of birth or interview and, therefore, cannot replicate these numbers exactly.

Locating the father was difficult for some cases, and in a few cases, the mother was interviewed after the father was interviewed. Therefore, before comparing mothers' and fathers' reports of time sensitive measures (i.e. relationship status, income), it is highly recommended data users check the time gap between parent interviews using the **cm1tdiff** constructed variable (available on the mother file).

At the time of the follow-up interviews, we attempted to interview the mother first. This was based on the assumption that, if the parents are not living together, the mother would be easier to locate and would have updated locating information about the father. There were, however, cases in which the mother was interviewed after the father. Mothers and fathers were also interviewed up to 14 months apart at one-year and 12 months apart at three-year. However, two-thirds were

---

[5] 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.

interviewed within one month of each other. Before comparing mothers' and fathers' reports of time sensitive measures (i.e. relationship status, income), data users should check the time gap between parent interviews using the **cm2tdiff/cm3tdiff/cm4tdiff** constructed variables (available on the mother files).

At the one-year follow-up, when asked about their relationship with the child's father at baseline (M2A6), 19 mothers responded that the father was a different father than the one provided by the interviewer. We were unable to determine for these cases whether these were truly different fathers or simply a minor coding error. Please use caution when analyzing these cases.

While each parent is asked to report on the focal child for many measures, some of the information on the child can only be obtained on the mother record. If you are examining father child relationships, you will need to use the mother data to obtain some child measures (such as child sex, birth weight, etc).

V. VARIABLES AND CODING

A. Variable Names
All raw variables begin with either "m" for mother or "f" for father. A "1" follows to indicate that the data is from the baseline interview. A "2" follows to indicate that the data is from the one-year follow-up, "3" that the data is from the three-year follow-up, and a "4" that the data is from the five-year follow-up. Variable names that begin with the letter "c" are constructed variables. The "c" is followed by either "m (1, 2, 3, 4)" or "f (1, 2, 3, 4)" for mothers or fathers, respectively. Note: constructed variables are not included in the electronic version of the questionnaire (see section below on constructed variables). Variable names starting with the prefix "mx" or "fx" were asked in the first two cities only.

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

All variables have labels and formats. 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:

       (-1) = Refused
       (-2) = Don't know
       (-3) = Missing
       (-4) = Multiple answers
       (-5) = Not asked (not in survey version)
       (-6) = Skipped
       (-7) = N/A
       (-8) = Out-of-range
       (-9) = Not in wave

Occasionally other codes were used (-10 or -14) to indicate the question did not apply to the respondent or the respondent had effectively provided a response via an earlier question.

B. Constructed Variables

A number of variables were constructed and added to the data set. Some represent data not otherwise available to the public, and some are merely aggregations of existing data that we provided 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.

Below we provide a table of constructed variables available in the file and a description of how we created some of the constructed variables (if the construction is not transparent). **Note:** Raw yes/no questions are coded as 1=Yes and 2=No. Constructed yes/no variables are coded as 1=Yes and 0=No.

Mother				Father			
Baseline	**One-Year**	**Three-Year**	**Five-Year**	**Baseline**	**One-Year**	**Three-Year**	**Five-Year**
Interview Flags							
cm1twoc	cm2twoc	cm3twoc	cm4twoc	cf1twoc	cf2twoc	cf3twoc	cf4twoc
	cm2mint	cm3mint	cm4mint	cf1fint	cf2fint	cf3fint	cf4fint
cm1fint	cm2fint	cm3fint	cm4fint		cf2mint	cf3mint	cf4mint
cm1tdiff	cm2tdiff	cm3tdiff	cm4tdiff				
cm1fdiff	cm2fdiff	cm3fdiff					
					cf2new12	cf3new30	cf4new60
		cm3inhom					
		cm3inccprov					
Sample Flags							
	cm2samp	cm3samp	cm4samp		cf2samp	cf3samp	cf4samp
cm1natsm	cm2natsm	cm3natsm	cm4natsm	cf1natsm	cf2natsm	cf3natsm	cf4natsm
cm1natsmx	cm2natsmx	cm3natsmx	cm4natsmx	cf1natsmx	cf2natsmx	cf3natsmx	cf4natsmx
cm1citsm	cm2citsm	cm3citsm	cm4citsm	cf1citsm	cf2citsm	cf3citsm	cf4citsm
Parents and Children							
cm1age	cm2age	cm3age	cm4age	cf1age	cf2age	cf3age	cf4age
cm1b_age	cm2b_age	cm3b_age	cm4b_age	cf1b_age	cf2b_age	cf3b_age	cf4b_age
cm1bsex							
cm1lbw							
cm1numb							
Relationships							
cm1relf	cm2relf	cm3relf	cm4relf				
cm1marf	cm2marf	cm3marf	cm4marf	cf1marm	cf2marm	cf3marm	cf4marm
	cm2amrf	cm3amrf	cm4amrf				
cm1cohf	cm2cohf	cm3cohf	cm4cohf	cf1cohm	cf2cohm	cf3cohm	cf4cohm
	cm2alvf	cm3alvf	cm4alvf				
	cm2finst						
	cm2stflg						

			Incarceration				
cm1finjail	cm2finjail	cm3finjail	cm4finjail	cf1finjail	cf2finjail	cf3finjail	
	cm2fevjail	cm3fevjail	cm4fevjail		cf2fevjail	cf3fevjail	
				cmf1finjail	cmf2finjail	cmf3finjail	
					cmf2fevjail	cmf3fevjail	
			Current Partner				
	cm2marp	cm3marp	cm4marp		cf2marp	cf3marp	cf4marp
	cm2cohp	cm3cohp	cm4cohp		cf2cohp	cf3cohp	cf4cohp
			Demographics and Household Composition				
cm1adult	cm2adult	cm3adult	cm4adult	cf1adult	cf2adult	cf3adult	cf4adult
cm1kids	cm2kids	cm3kids	cm4kids	cf1kids	cf2kids	cf3kids	cf4kids
cm1gdad	cm2gdad	cm3gdad	cm4gdad	cf1gdad	cf2gdad	cf3gdad	cf4gdad
cm1gmom	cm2gmom	cm3gmom	cm4gmom	cf1gmom	cf2gmom	cf3gmom	cf4gmom
cm1edu				cf1edu			
cm1ethrace				cf1ethrace			
	cm2biok						
	cm2fbir						
			CIDI				
		cm3alc_case				cf3alc_case	
		cm3drug_case				cf3drug_case	
	cm2gad_case	cm3gad_case			cf2gad_case	cf3gad_case	
	cm2md_case_con	cm3md_case_con	cm4md_case_con		cf2md_case_con	cf3md_case_con	cf4md_case_con
	cm2md_case_lib	cm3md_case_lib	cm4md_case_lib		cf2md_case_lib	cf3md_case_lib	cf4md_case_lib

			Household Income and Poverty				
cm1hhinc	cm2hhinc	cm3hhinc	cm4hhinc	cf1hhinc	cf2hhinc	cf3hhinc	cf4hhinc
cm1hhimp	cm2hhimp	cm3hhimp	cm4hhimp	cf1hhimp	cf2hhimp	cf3hhimp	cf4hhimp
					cf2hhincb	cf3hhincb	cf4hhincb
					cf2hhimpb	cf3hhimpb	cf4hhimpb
cm1inpov	cm2povco	cm3povco	cm4povco	cf1inpov	cf2povco	cf3povco	cf4povco
cm1povca	cm2povca	cm3povca	cm4povca	cf1povca	cf2povca	cf3povca	cf4povca
					cf2povcob	cf3povcob	cf4povcob
					cf2povcab	cf3povcab	cf4povcab
			Cognitive Ability				
		cm3cogsc				cf3cogsc	
			Interview Type				
	cm2tele	cm3tele	cm4tele	cf1tele	cf2tele	cf3tele	cf4tele
cm1span	cm2span	cm3span	cm4span	cf1span	cf2span	cf3span	cf4span

Constructed measure definitions and/or notes about constructions

Child age at interview
(cm*b_age, cf*b_age)
Scale changes across waves. Baseline variables (cm1b_age, cf1b_age) are measured in days while one-, three-, and five-year child age measures are constructed in months.

Mother relationship with father
Response categories change between baseline (cm1relf) and one-year (cm2relf); three-year (cm3relf) and five-year (cm4relf) response categories are the same as one-year.

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.

Mother's reported romantic relationship with baby's father at one-, three-, and five-year
(cm2relf, cm3relf, cm4relf)
In the one-year and three-year follow-up, the relationship status variable is based on information reported by a mother about her relationship status with the baby's father (m2a7/m3a4/m4a4), and cohabitation status as reported in question (m2a7a/m3a4a1/m4a4a1).

Mothers are considered married to the focal child's father for cm2relf/cm3relf/cm4relf if (m2a7/m3a4/m4a4=1). For mothers who report to be romantically involved (m2a7/m3a4/m4a4=2), m2a7a/m3a4a1/m4a4a1 is tabulated to determine the cohabitation status. Mothers who are romantically involved and live with their respective babies' fathers "all or most of the time" are considered to be romantically involved – cohabiting (cm*relf=2). Mothers who are romantically involved with the respective babies' fathers but live with father only "some of the time" are coded as rom-some visit (cm*relf=3). Mothers who are romantically involved with the respective babies' fathers but live with them only "rarely", "never" or "rarely/never" are coded as rom-no-visit (cm*relf=4). Mothers who don't live with the respective babies' fathers due to separation, divorce or death are coded as "sep/div/wid" (cm*relf=5). The three additional categories in the cm*relf variable: "friends", "not in any kind of relationship" and "father unknown" are based on mothers' report in m2a7/ m3a4. Two specific cases in the one-year follow-up reporting romantic involvement in m2a7, but unsure whether cohabiting or not are coded as missing (-3). Four cases where mother reported "father unknown" but we have father interviews were recoded based on father reports after confirming key facts about the couple.

Low birth weight
(cm1lbw)
Babies weighing less than 2,500 grams at birth are defined as low-birth-weight babies. The variable (cm1lbw) is coded one (1) for any baby weighing less than 2,500 grams at birth, and zero (0) for babies that 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.

Cognitive ability
(cm3cogsc, cf3cogsc)
Cognitive ability is measured as the sum of the correct items in the Similarities subtest of the Wechsler Adult Intelligence Scale – Revised (WAIS-R). See the three-year scales documentation for more details on this measure. NOTE: For some fathers, this measure was administered at the one-year follow-up. Therefore there may be a cognitive score on the three-year record for fathers not interviewed in the three-year follow-up.

Father lives in the same state as the mother
(cm2finst, cm2stflg)
This question was not asked in the first two cities. To fill in the missing data, we used the three-year data if available. If we did not have a mother's report of father's state of residence at one-year we checked state of residence at the three-year follow-up. Couples who reported to be living together at the one- and three-year follow-ups, were assumed to be cohabiting continuously and coded (1) on CM2FINST. For the remaining cases, if the father was reported to be living in the same state as the mother at the three-year follow-up, it was assumed that he did not move out and back in to the state and was coded (1) in state at one-year. If father was reported to be living in a different state than mother at the three-year follow-up (1%) we assumed he lived out of state at the one-year follow-up and coded (0) on cm2finst. Cases that are still missing information are coded missing (-3). The flag cm2stflg indicates the imputed data.

More detailed information about the geographical distance between parent's places of residence will be collected at the five-year follow-up survey. It will include a question in which the respondent will be asked to report how far apart the other parent lives from him/her. The respondent will select from a predetermined range, given in miles.

The following are constructed variables that are new or have changed significantly between the 2005 and 2008 versions of the first three waves.

Father in jail
(cf*finjail, cmf*finjail, cf*fevjail, cmf*fevjail)
The constructed jail variables for mother report of father in jail, father report of his own jail, combined reports, and cumulative measures of whether father has ever been in jail are available at each wave. 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.

Parents' education
(cm1edu, cf1edu)
Constructed variables for mothers' and fathers' education at baseline; mothers' and fathers' 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.

Parents' race/ethnicity
(cm1ethrace, cf1ethrace)
Constructed mothers' and fathers' race/ethnicity variables; mothers' and fathers' are based their own reports, but mothers' reports of fathers' race are used for fathers who were not interviewed at any wave.

CIDI alcohol and drug, depression, and generalized anxiety
Alcohol and drug abuse, depression (both liberal and conservative measures), and generalized anxiety disorder based on the CIDI-SF diagnostic are available at some follow-up interviews (see constructed variable measures chart for what is available at each wave)Please see the Scales Documentation for more information on the CIDI scales implemented the Fragile Families Study.

Household income and poverty
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, employed last year, hours worked, total adults in household, earnings, received welfare, and

marital status. For couples that 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.

At each follow-up interview (**cm2hhinc, cf2hhinc, cm3hhinc, cf3hhinc, cm4hhinc, cf4hhinc**), respondents were asked to provide an exact dollar amount. If they could not, they were asked to provide a range. This strategy was effective in reducing missing data to about 10 percent, although a portion of parents reported a range rather than an exact dollar amount. To construct household income at the follow-up waves, we first imputed dollar amounts for those who reported a range of income (using others who provided income in the same range but provided a detailed amount of income). Next, we imputed dollar amounts for those with no reported income. Both imputations included the following covariates: relationship status (mother report), age, race/ethnicity, immigrant, employed last year, earnings, total adults in the household, and received welfare. Imputations for those who reported a range were based on parent's own characteristics. Imputations for missing income were based on both parent's characteristics for married and cohabiting couples; otherwise, they were based on parent's own characteristics. An additional set of father variables (**cf2hhincb, cf3hhincb, cf4hhincb**) were created using mother reports of household income for married and cohabiting couples. These variables are comparable to how the baseline father household income variable was created.

A series of imputation flags (**cm1hhmip, cf1hhimp, cm2hhimp, cf2hhimp, cf2hhimpb, cm3hhimp, cf3hhimp, cf3hhimpb, cm4hhimp, cf4hhimp, cf4hhimpb**) 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 in these flags. Users can examine the raw variables to determine who had detailed/bracketed data. Note: 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.

Poverty measures suffixed with "povco (inpov at baseline)" are the ratio of total household income (as defined in the variables described above) to the official poverty thresholds established by the U.S. Census Bureau. The poverty measures suffixed with "povca" transform the ratios into categorical variables. The thresholds 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 "b" versions of the poverty variables for fathers are based on the "b" versions of his household income variables. The imputation flags created for the household income variables also refer to the poverty variables.

Please visit https://www.census.gov/hhes/www/poverty/data/threshld/ for detailed information about poverty thresholds.

C. Data Cleaning
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.

D. Open-ended response codes
Free response questions (open-ended questions) were coded by CRCW staff. Codes were assigned by two CRCW 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 are recoded into new categories in the 100 range (101, 102, etc). Cases that indicate an "other" but were vague or unique remain coded simply as "Other (not specified)."

Occupations
We constructed an occupation variable for each respondent at each wave (except for mothers at baseline when the question was not asked) 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 (Group A)
> 102 - Executive, Administrative, and Managerial Occupations (Group B)
> 103 - Sales Occupations (Group C)
> 104 - Administrative Support Occupations, including Clerical (Group D)
> 105 - Precision Production, Craft, and Repair Occupations (Group E)
> 106 - Machine Operators, Assemblers, and Inspectors (Group F)
> 107 - Transportation and Material Moving Occupations (Group G)
> 108 - Handlers, Equipment Cleaners, Helpers, and Laborers (Group H)
> 109 - Service Occupations, except Private Household (Group K)
> 110 - Unspecified
> 112 - Military
> 113 - Farming/Agriculture (*father baseline only)*
> 114 - Self-employed (*father baseline only)*

Occupations are located in variables f1J7bc, m2k10bc, f2k15bc, m3k13, and f3k12 and are based on job titles and duties in *regular* employment.

Codes for occupations in "other" types of work (e.g. baseline **FJ13B** – work in own business and **FJ13D** – other source of income), were coded using a slightly different set of categories designed by CRCW staff that incorporated some additional categories necessitated by the data. When appropriate, CRCW staff followed the classifications described by *Occupational Classification System* by Major Occupational Groups (though these code numbers differ slightly).

101　ARTISTS AND ATHLETES – to include 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 & PRECISION TRADES – to include 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 – to include escort service, adult entertainment, party services, DJs, and gambling.

107　TRANSPORTATION & DELIVERY

108　SERVICE OCCUPATIONS – to include 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 & REPAIRERS – to include 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 & FINANCE

112　LANDSCAPING & AGRICULTURE – to include landscaping, cutting grass, ranching, farming, raising cattle.

113　PROFESSIONAL – to include 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.

VI.　SAMPLE 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, nonmarital births were oversampled. Therefore, Mathematica Policy Research has 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 Mother, Father, and Couple Replicate Weights for Core Telephone Surveys Waves 1-4" For detailed information on the construction of the sample weights, please read "Fragile Families &

VII.   DATA COLLECTION PROTOCOLS

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 included in the study. At baseline, eligible mothers are asked to identify the father of the child, and fathers are interviewed in person during hospital visits or by telephone. The one-year, three-, and five-year follow-up interviews were designed to be conducted by telephone using a Computer Assisted Telephone Instrument (CATI). Cases where parents cannot be located or interviewed by telephone are sent to the field.   Field interviewers trained to administer the collaborative In-Home Longitudinal Study of Pre-School Aged Children were also used to locate and interview mothers and fathers.

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. Approximately 30 percent of mothers and 25 percent of fathers interviews at one-year were conducted over the phone; the remaining interviews were conducted in-person.   About 98 percent of mothers and 95 percent of fathers were interviewed by phone at the three and five-year follow-up surveys.

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 2001 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.

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)[6] 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

---

[6] 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 and to serve as the survey subcontractor for subsequent rounds of data collection in all 20 cities.

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).[7] See [Sample Sizes and Response Rates](#) in Section III for mothers' and fathers' response rates at baseline and each follow-up.

All respondents who completed a baseline interview were contacted for the one-, three-, and five-year follow-up surveys, as were non-respondent at baseline fathers whose partner (mother) had completed a baseline interview. A small portion of the original respondents was found to be ineligible at the time of the follow-up interviews. See the sample flags (c*samp) for counts at each wave. Reasons for considering a case ineligible for further interview include: parent deceased, child deceased, child adopted, (and for fathers) DNA confirmation that the original respondent is not the child's father.

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.

All mothers who completed a baseline interview and who remained eligible were contacted for each follow-up interview.   All follow-up mother interviews were first attempted by telephone using CATI. In cases in which we could not contact the mother by telephone, local field interviewers were assigned cases requiring field locating. The field interviewers were encouraged to have respondents call a 24-hour toll-free number at the MPR survey operations center to complete the interview on the CATI system. Field interviewers were also trained in administration of the survey instrument. Respondents completing the one-year, three-year, and five-year interviews by telephone were provided with $30 incentive payment. Those requiring a field visit to complete the core survey were provided with $50 incentive payment.

---

[7] Some father interviews took place before the mother interview was completed. If a mother was not eventually completed, the father was dropped from the sample, however.

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. Father follow-up interviews followed the same protocols and incentives as mothers.

Some fathers were incarcerated at the time of data collection in their location. In these cases, MPR staff worked to obtain special clearance, including permission from the Federal Bureau of Prisons, to conduct interviews with incarcerated respondents. When possible for cost containment purposes, interviews with incarcerated respondents were attempted by telephone. However, some prisons do not permit telephone interviews. In those cases MPR field interviewers arranged for in-person visits.

VIII. APPENDIX A. SAMPLE ELIGIBILITY SCREENING FORM

ID#: |__|__|__|__|__|__|__|__|__| INT. ID #: _____ DATE: _____ STATUS: ☐ Eligible ☐ Ineligible

SURVEY OF NEW PARENTS: BASELINE SCREENER

Hello, I'm (YOUR NAME) from Mathematica Policy Research (SHOW CARD). I am here to talk with you about an important survey we are conducting with new parents. First, I would like to congratulate you on the birth of your new baby/babies. Now let me tell you a little bit about the survey. (HAND STUDY BROCHURE). Have you seen our brochure? This study, which is sponsored by a variety of well-known foundations, will help improve researchers' and policymakers' understanding of the needs and concerns of new parents. Do you have a few minutes to talk with me now?

1. First, what is your baby's name?

   INTERVIEWER: IF MULTIPLE BIRTHS, RECORD NAMES IN BIRTH ORDER.

   _____

   -4 ☐ Not named yet

1a. When was [BABY] born?

   |__|__| / |__|__| / |__|__|__|__|
   MONTH   DAY   YEAR

2. Are you currently married to [BABY'S] father?

   1 ☐ YES → GO TO 2a
   0 ☐ NO → SKIP TO 2c

2a. **INTERVIEWER: HAS YOUR FIELD MANAGER INSTRUCTED YOU TO STOP INTERVIEWING MARRIED COUPLES?**

   1 ☐ YES → GO TO 2b
   0 ☐ NO → SKIP TO 2c

2b. **INTERVIEWER READ TO R:** Unfortunately, we are only interviewing single parents today, but thank you for taking the time to talk with me. Best of luck with your new baby! **END INTERVIEW.**

2c. Is [BABY'S] father still living?

   1 ☐ YES → GO TO 3
   0 ☐ NO → SKIP TO 9
   d ☐ DON'T KNOW/DON'T KNOW FATHER → GO TO 3

3. How old are you? |__|__|

3a. **INTERVIEWER: IS R UNDER 18?**

   1 ☐ YES → GO TO 3b
   0 ☐ NO → SKIP TO 4

3b. **INTERVIEWER: DOES THIS HOSPITAL ALLOW YOU TO INTERVIEW R UNDER THE AGE OF 18?**

   1 ☐ YES → GO TO 4
   0 ☐ NO → SKIP TO 3d

   INTERVIEWER READ TO R: We need to interview parents who are 18 years old or older, so unfortunately, we won't be able to interview you. But thank you for taking the time to talk with me, and best of luck with your new baby. **END INTERVIEW**

3c. **INTERVIEWER: IF 2c ="d", CODE "d" FOR 3d AND SKIP TO 9.**

3d. How old is [BABY'S] father? |__|__|

   -1 ☐ DON'T KNOW/ OLDER THAN 18 ⟶ SKIP TO 5
   -2 ☐ DON'T KNOW/ YOUNGER THAN 18 ⟶ SKIP TO 4d
   d ☐ DON'T KNOW ⟶ SKIP TO 5

4a. What is his date of birth?

   |__|__| / |__|__| / |__|__|__|__|
   MONTH   DAY   YEAR

   d ☐ DON'T KNOW
   r ☐ REFUSED

4b. **INTERVIEWER: IS FATHER UNDER 18?**

   1 ☐ YES → GO TO 4c
   0 ☐ NO → SKIP TO 5

4c. **INTERVIEWER: DOES THIS HOSPITAL ALLOW YOU TO INTERVIEW R UNDER THE AGE OF 18?**

   1 ☐ YES → SKIP TO 5
   0 ☐ NO → GO TO 4d

4d. **INTERVIEWER READ TO R:** We need to interview parents who are both 18 years old or older, so unfortunately, we will not be able to interview you. But thank you for taking the time to talk with me, and best of luck with your new baby! **END INTERVIEW**

C Will your baby live with you?

   ₁ ☐ YES → **SKIP TO 6**

   ₀ ☐ NO → **GO TO 5a**

5a. Who will the baby live with?

   ₃ ☐ BABY'S FATHER ⟶ **SKIP TO 7**

   ₂ ☐ MOTHER'S RELATIVE(S) ⟶ **SKIP TO 7**

   ₁ ☐ FATHER'S RELATIVES ⟶ **SKIP TO 7**

   ₀ ☐ ANY OTHER RESPONSES ⟶ **GO TO 6**

6. **ASK, IF NOT VOLUNTEERED:** Are you planning on giving your baby up for adoption?

   ₁ ☐ YES → **GO TO 6a**

   ₀ ☐ NO → **SKIP TO 7**

6a. **INTERVIEWER READ TO R:** Since you are planning on giving your baby up for adoption, we will not need to bother you with any more questions. Thank you for taking the time to talk. **END INTERVIEW**

7. Before we begin the interview, please tell me how you spell your name (and your baby's father's name).

A. MOTHER'S NAME:

FIRST	LAST

   ᵣ ☐ REFUSED

B. FATHER'S NAME:

FIRST	LAST

               SKIP TO 8

   d ☐ DON'T KNOW → **GO TO 7a**

   ᵣ ☐ REFUSED → **SKIP TO 8**

7a. And why is that?

DO NOT READ

   ₄ ☐ BABY RESULT OF FORCED SEX ⟶ **SKIP TO 9**

   ₃ ☐ ONE NIGHT STAND ⟶ **SKIP TO 11**

   ₂ ☐ FATHER UNKNOWN ⟶ **SKIP TO 11**

   ₁ ☐ DONOR INSEMINATION ⟶ **SKIP TO 11**

   ₀ ☐ OTHER (SPECIFY) ⟶ **SKIP TO 11**

_____

8. Is [BABY'S] father planning to come to the hospital today or tomorrow?

   ₃ ☐ YES, HE'S HERE NOW ⟶ **SKIP TO 10**

   ₂ ☐ YES, COMING LATER TODAY → **GO TO 8a**

   ₁ ☐ YES, COMING TOMORROW → **GO TO 8a**

   ₀ ☐ NO, NOT COMING ⟶ **SKIP TO 8b**

   -₁ ☐ FATHER DOESN'T KNOW ABOUT BABY ⟶ **SKIP TO 11**

   d ☐ DON'T KNOW IF HE WILL COME ⟶ **SKIP TO 11**

8a. What time is he expected?

|___|___|:|___|___| AM PM

       SKIP TO 11

8b. Is there any particular reason why he isn't coming?

DO NOT READ

   ₆ ☐ DECEASED ⟶ **GO TO 9**

   ₅ ☐ BABY RESULT OF FORCED SEX ⟶ **GO TO 9**

   ₄ ☐ DOESN'T KNOW ABOUT BABY ⟶ **SKIP TO 11**

   ₃ ☐ MOTHER AND FATHER DON'T TALK OR SEE EACH OTHER ⟶ **SKIP TO 11**

   ₂ ☐ FATHER UNKNOWN ⟶ **SKIP TO 11**

   ₁ ☐ NO REASON/OTHER (SPECIFY) ⟶ **SKIP TO 11**

_____

_____

9. **INTERVIEWER READ TO R:** We need to interview both mothers and fathers, so unfortunately, we won't be able to interview you for this study. But thank you for taking the time to talk with me! **END INTERVIEW.**

10. **INTERVIEWER READ TO R IF FATHER IS AVAILABLE:** While [BABY'S] father is here, I'd like to go ahead and interview him. I will come back and complete the survey with you when we are finished. Thank you for being flexible.

11. **INTERVIEWER READ TO R:** Now I would like to ask you to participate in the Survey of New Parents. Please read and sign this consent form. I will be happy to answer any questions you might have.

INTERVIEWER: NOW BEGIN QUESTIONNAIRE WITH MOTHER.

IX. APPENDIX B. CHANGES TO THE FIRST THREE WAVES OF THE CORE DATA FILES BETWEEN 2005 AND 2008

Data File Updates
The table below shows the file names for the current public release and the corresponding file each is intended to replace. All data users are required to update their files to the new versions. We also encourage users to use the merged file which contains all core data available to the public.

	Current Release		Retired	
	Mother	Father	Mother	Father
Core				
Baseline	ffmombspv3	ffdadbspv3	ffmombspv2	ffdadbspv2
One-Year	ffmom1ypv2	ffdad1ypv2	ffmom1yp	ffdad1yp
Three-Year	ffmom3ypv2	ffdad3ypv2	ffmom3ypv1	ffdad3ypv1
Five-Year	ffmom5ypv1	ffdad5ypv1	N/A	N/A
Merged	ff_pub_merge2		N/A	

Weights and Sampling
The final version of sample weights for the national sample and full 20 cities sample are now available to the public. This includes sets of replicate weights for estimating variance in lieu of stratum and PSU variables (available only via restricted use contract). For a brief introduction to using the weights and a complete discussion on how the weights were designed, please see the documents linked in Section VI. Sample Weights.

The sample flags have also been revised as part of the weights design process. The national sample flags (cm*natsm, cf*natsm) have changed to reflect the randomly selected sample in the national weights. Two new sets of sample flags (cm*natsmx, cf*natsmx) and (cm*citsm, cf*citsm) have been added to each wave to reflect the sample sizes of cases in the national sample minus one city and the 20 cities sample. The constructed variables that indicate the reasons for non-response (cm*samp, cf*samp) have been revised to break out the non-response into even more detail (including two new categories: Refusal and Could Not Locate.)

Case Status Changes
One mother and her father requested to be removed from the sample and are now coded as "Not in wave" at all waves. The only complete interview where valid data was overwritten was the mother's baseline interview.

One mother and her father have been overwritten and coded as "Not in wave" at the one-year follow-up interviews because they did not report on the focal child.

One father that should not have been followed and interviewed at the one-year follow-up has been overwritten and coded as "Not in wave."

Different Dads
Some mothers provided conflicting information over the waves about who is the biological father of the child. The cm*fdiff variables flag such cases where the mother indicated the biological father was a different man than had been indicated at earlier waves and for whom we had no reason to doubt this information. However, we cannot determine the accuracy of these reports.

Constructed Variables

New constructed variables (many not previously available to the public) have been added to the files in addition to revisions to existing constructed variables. These include:

- Sample flags
- Two cities flags
- Ages, Education, Race/Ethnicity
- Child gender, Different dad flags
- Incarceration
- CIDI
- Household income and poverty

For a complete discussion of additions/changes, please revisit B. Constructed Variables in Section V.

Data Cleaning

Extensive cleaning was limited to mothers' and fathers' one-year follow-up interviews particularly in Sections, C, G, and K. The primary objective was to resolve skip pattern issues between two cities and 18 cities respondents. No interview data was overwritten during this process, but we were able to recover some data for the following questions: m2c19a, m2g5b, m2k13a, m2k13b, m2k13c, f2g7b, f2k5. In addition, 14 two cities cases have been recovered and added to the fathers' one-year file.

Cleaning birth dates resulted in minor changes to the constructed age variables from previous versions of the files.

15 cases with missing data on child gender have been recovered and 22 cases gender have been changed to reflect additional we gained from the medical records data.

A one case discrepancy between the constructed relationship variable (cm1relf) and this variable has been reconciled.

The constructed baseline cohabitation variables no longer include parents who are married. This construction matches how the corresponding variables are created at follow-up waves.

X.   KNOWN ISSUES

This section highlights known issues and errors in the Fragile Families data sets that could not be cleaned or where data could not be recovered. Users should review this information to plan their analysis accordingly.

Non-custodial Fathers and Child Support (Father Three-Year): Question C1C, that asks if mother has primary custody of child, should have asked if father has primary custody. This resulted in non-custodial fathers skipping the child support question in section C they should have answered.

Smoking and Drinking (Three-Year): In both the mother and father surveys, a large number of cases skipped questions about smoking and drinking (J31-J34) due to an error in the CATI program. Information for these cases could not be recovered.

Kindergarten (Five-Year): The questions that ask if schools are on summer break (B7) and if child is currently enrolled in kindergarten (B7A) are not reliable indicators. Interviewers were instructed to code these questions without asking. The results in B7 do not line up consistently with the interview date, and a large number of cases in B7A are coded as skip without any skip instruction. Only respondents in the two cities have valid data on kindergarten enrollment (B8A). We recommend that users not use questions B7 and B7A for determining if child is currently enrolled in kindergarten.