

Fragile Families & Child Wellbeing Study: A Brief Guide to Using the Mother, Father, and Couple Replicate Weights for Core Telephone Surveys Waves 1-4

April 2008

This memo will provide a brief overview of the weights for the Fragile Families core telephone data and how to implement them using Stata.ⁱ All data users should review this memo, including the **endnotes**, in order to understand the different weights on the file, how to use them, and what they imply for your estimates. A separate memo “*Fragile Families & Child Wellbeing Study: Methodology for Constructing Mother, Father, and Couple Weights for Core Telephone Surveys Waves 1-4*” explains how the core mother, father, and couple weights were created. Users should also review Reichman et al 2001ⁱⁱ for details on the sample design. This memo is not intended to provide statistical advice. It simply describes the weights that are available to data users and other key information needed to understand and apply these weights. Statistical advice should be sought from other sources (resources at your institutions, statistical programming manuals, etc.).

Using weights with the Fragile Families data is important in order to get unbiased statistical estimates. The 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. All of the weights described in this memo adjust for the sample design (probability of selection), non-response at baseline, and attrition based on observed characteristics over the waves (see the Weights Methodology memo for more details). Unweighted estimates may yield misleading results.

The main purpose of this is memo is to describe the weights available to users of the public data files. The public use data do not contain the geographic identifiers needed to construct the strata and primary sampling unit (PSU) variables necessary for using a Taylor Series methodology to estimate variances (except through a restricted use contract)ⁱⁱⁱ. 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 strata 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. Available weights^{iv}

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^v 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^{vi} 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.

When setting the weights in your statistical package, you will need to indicate the name of the basic weight, as well as the replicate weights. The variable names for these weights are listed below where (m) indicates mother weights, (f) indicates father weights, and (c) indicates couples weights and 1-4 represents the wave (1 = baseline, 2 = one-year follow-up, 3 = three-year follow-up, 4=five-year follow-up).

Wave	National level weights ^{vii}		City level weights	
	Basic weight	Replicate weights	Basic weight	Replicate weights
Baseline ^{viii}	m1natwt f1natwt	m1natwt_rep1-m1natwt_rep33 f1natwt_rep1-f1natwt_rep33	m1citywt f1citywt	m1citywt_rep1- m1citywt_rep10 f1citywt_rep1- f1citywt_rep10
One-Year	m2natwt f2natwt c2natwt	m2natwt_rep1-m2natwt_rep33 f2natwt_rep1-f2natwt_rep33 c2natwt_rep1-c2natwt_rep33	m2citywt f2citywt c2citywt	m2citywt_rep1- m2citywt_rep10 f2citywt_rep1- f2citywt_rep10 c2citywt_rep1- c2citywt_rep10
Three-year	m3natwt f3natwt c3natwt	m3natwt_rep1-m3natwt_rep33 f3natwt_rep1-f3natwt_rep33 c3natwt_rep1-c3natwt_rep33	m3citywt f3citywt c3citywt	m3citywt_rep1- m3citywt_rep10 f3citywt_rep1- f3citywt_rep10 c3citywt_rep1- c3citywt_rep10
Five-year	m4natwt f4natwt c4natwt	m4natwt_rep1-m4natwt_rep33 f4natwt_rep1-f4natwt_rep33 c4natwt_rep1-c4natwt_rep33	m4citywt f4citywt c4citywt	m4citywt_rep1- m4citywt_rep10 f4citywt_rep1- f4citywt_rep10 c4citywt_rep1- c4citywt_rep10

As an example, to weight the baseline mother data to be representative of births in the 77 cities, you would apply the weight **m1natwt** (and replicate weights **m1natwt_rep1- m1natwt_rep33**). If you were looking at a father-reported measure at the three-year follow-up and wanted to use all 20 cities but not generalize beyond those 20 cities, you would apply the weight **f3citywt** (and replicate weights **f3citywt_rep1- f3citywt_rep10**). If you were analyzing the sample of couples at the three-year follow-up and wanted to compare reports of father involvement among a nationally-representative sample, you would apply the **c3natwt** (and replicate weights **c3natwt_rep1- c3natwt_rep33**).

These weights are cross-sectional in nature and each weight makes its particular wave representative of the original sampling frame. Longitudinal weights, which would make cases interviewed at every wave representative of the original sampling frame, are not available. Therefore, if you are using measures from more than one wave, we suggest using the weight of the wave in which the most people were interviewed.

B. Applying the weights and computing variance estimates using Stata (version 9 or later)

As described in the weights construction memo, the replicate weights require using jackknife estimation of standard errors. The basic weight is a sampling weight (which means that the weight implies the number of people in the population represented by the sampled individual) and therefore should be set using the `pweight` option in Stata (see the survey data manual in stata). The syntax of the `svyset` statement you should use is ...

svyset [pweight=BASICWEIGHT], jkrw(REPLICATES, multiplier(1)) vce(jack) mse

... where *BASICWEIGHT* and *REPLICATES* are replaced with the relevant weight names for your analyses.

For instance, to run an analysis of baseline mother data for the national sample, the data user should first set the weight by entering...

```
svyset [pweight=m1natwt], jkrw(m1natwt_rep*, multiplier(1)) vce(jack) mse
```

To run an analysis of three-year father data for the 20-city sample, the data user should first set the weight by entering...

```
svyset [pweight=f3citywt], jkrw(f3citywt_rep*, multiplier(1)) vce(jack) mse
```

Then simply add “svy:” before (nearly) any of their commands (e.g. tab, reg, or logistic) to generate weighted estimates.

If you are running on a subgroup of the total population, you also need to include the subpop command with the appropriate sample flag...

```
svy, subpop(SAMPLEFLAG):
```

...before the command you are running.

For instance, to run an analysis of baseline mother data for the national sample, the regression command would look like ...

```
svy, subpop(cm1natsm): regress Y X
```

Using the subpop option in Stata is preferred over subsetting the file (e.g. with an if or keep statement) to the domain of interest. By including the full file and specifying the subpopulation command, Stata will make full use of the design when determining the degrees of freedom.

For more information on using weights in Stata, users should review the Stata manuals, particularly the “Survey Data” manual.

C. Identifying the national and city-level samples

There is a set of sample flags that correspond to the weights that identify the sample members who will be included in your weighted estimates. You can use these variables to select the sample to run unweighted estimates on for comparison to their weighted estimates or in the subpop option when running regressions. We generated these because there are a small number of cases that do not have weights but have valid survey data (see endnote v for more detail) and because 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 for more information).

The weights sample flags are dummy variables that are “1” if the individual/couple is in that sample and has a valid interview(s) and “0” if the individual/couple is interviewed but not in that sample and are “.” (missing) if the individual/couple was not interviewed

in that wave. These flags can be used to subset out the population of interest (see use of the subpop command in Appendix B). The table below displays the name of the weight, the corresponding sample flag, and a description of the sample implied by the flag (where Z is the indicator for mother (m), father (f), or couple (c) and W is the indicator for wave – 1 = baseline, 2 = one-year follow-up, 3 = three-year follow-up, 4=five-year follow-up).

Weight	Sample flag	You can use this flag to select these cases
ZWnatwt	cZWnatism	Randomly selected respondents in the randomly selected 16 cities that have a completed core interview
ZWcitywt	cZWcitsm	Randomly selected respondents in the 20 cities who have a completed core interview [See endnote v for note re: cases dropped]

For instance, cm4natism = 1 indicates that the mother is in the national sample and interviewed at the five-year follow up. And cc3citsm = 1 would mean both mother and father were interviewed at the three-year follow-up and are in the 20-cities sample.

There are also flags (innatism and incitysm) that indicate whether the case is part of the national or city sample, but do not indicate whether the case was interviewed in any particular wave.

Sample sizes

The table below shows the sample sizes of completed interviews (with valid weights) in the national sample and 20-cities sample core data. Cases with deaths and adoptions are not included in these counts (unless the mother completed an interview, which occurred for some cases in which the child was living with neither parent). Whether a child/respondent has died or been adopted can be determined by the interview sample flag (ZWsamp) – see the *Guide to the Public Use Files* for more detail on the interview sample flags and appendix B for more information.

	Baseline	One-Year	Three-Year	Five-Year
Mother sample sizes				
National sample	3,442	3,082	2,973	2,927
City sample	4,789	4,270	4,140	4,055
Father sample sizes				
National sample	2,726	2,409	2,315	2,235
City sample	3,742	3,306	3,225	3,087
Couple sample sizes				
National sample	2,726	2,349	2,220	2,128
City sample	3,742	3,218	3,093	2,936

Appendix A - Changes to the Fragile Families Weights

Some users may be interested in how these weights (and sample flags) differ from previous versions. In addition to some methodological differences (e.g. changing the annualization factor), some practical differences that data users may recognize are:

- A handful of cases were dropped from national sample and city sample weights because the cases were not randomly selected as part of the core sample. These cases are coded as 0, or “No” in the city sample flags.
- We now provide city-level sample weights (for all 20 cities), father weights, and couple weights.
- We now include a weight to account for missing data in one of the national sample cities when a needed question was not asked in that city (see endnote vi).
- Data users can now estimate the proportion of children or parents who died and/or children living with neither parent (See appendix B).

Appendix B – Examining sample members who are deceased or who have deceased or adopted children

Sample members who have children who are deceased or adopted or are deceased themselves do not have survey data. Cases in which the child was living with neither parent may or may not have survey. Therefore, most of these cases will be excluded from analyses because of missing information. Cases with no interview data (just information on their status from dispositions) also have zeros in the weights sample flag. These cases do have valid weights, however, and the percent of individuals with these outcomes can be estimated by applying the weight to the sample flag. For instance, if you want to estimate the percent of fathers in the 20 cities who died by the three-year follow-up, the data user can tabulate the interview summary flag, in this case **cf3samp** (see the *Guide to the Public Use Files*), with weights. Note: cases reported as living with neither parent or even adopted may have valid interview data a later wave.

ⁱ This memo describes the weights that were released in February of 2008. See “Appendix A. Changes to the Fragile Families Weights” for a brief description about how these weights and sample flags differ from earlier versions.

ⁱⁱ Reichman, N. E., J. O. Teitler, I. Garfinkel, and S. S. McLanahan. 2001. "Fragile Families: Sample and Design." *Children and Youth Services Review* 23:303-326.

ⁱⁱⁱ 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.

^{iv} There are no weights that adjust for non-response to the in-home and child care studies. Users should treat this non-response as they would other item non-response.

^v In this memo, the term *national* refers to all 77 U.S. cities with 1994 populations of 200,000 or more.

^{vi} 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).

^{vii} There is also an alternate set of national sample weights with the suffix “x” (e.g, **m1natwtx**). Applying these weights makes the data from 15 of the cities in the sample representative of births occurring in large U.S. cities. This weight achieves the same goal as the primary national sample weight (described above) but drops one of the cities that has a high rate of questions “not asked” (denoted by -5 in the data), particularly at the one-year follow-up, because of changes to the survey instruments between fielding the first two cities and the remaining 18 cities. For example, this weight could be used if you wanted to analyze responses in the following variables... m2b11, m2b11a, m2b11b, m2b21, m2c10, m2d4, m2d4a, etc. To identify variables when the alternative national weights may be necessary, see the annotated questionnaires for notations “18-cities only” or see data for questions with high percent of “-5” responses.

^{viii} Couple weights are the same as father weights at the baseline interview (since all mothers were interviewed). For analyzing couples at baseline, the father baseline weight (f1natwt, f1natwtx, or f1citywt and corresponding replicate weights-depending on the sample/measure of interest) should be used. See page 28 of the “Fragile Families & Child Wellbeing Study: Methodology for Constructing Mother, Father, and Couple Weights for Core Telephone Public Survey Data Waves 1-4.”