THE FRAGILE FAMILIES AND CHILD WELLBEING STUDY DATASET: MULTIPLE IMPUTATION FOR THE MISSING FATHERS

Soma Roy, Department of Statistics, The Ohio State University
Sharon L. Sassler, Department of Sociology, The Ohio State University
Elizabeth A. Stasny, Department of Statistics, The Ohio State University

Abstract:

In recent times, the number of child births outside marriage has increased dramatically. As a result, a new form of family, called the "fragile family", has come into being. Mincy, 1994, and Mincy and Pouncy, 1997, define the "fragile family" as "unmarried parents who are raising their children together." These families are of great interest to policy makers and community leaders, because studies show that the wellbeing of children depends heavily on the relationship of their parents.

In this paper, we study data on fragile families from the Fragile Families and Child Wellbeing Study. These data were collected in the hospital shortly after the birth of a child. In about 22% of the cases the birth father was not present to be interviewed. We have evidence from previous studies that the fathers, who are missing at the time of, or shortly after the birth of the child, share some characteristics that are generally different from those of fathers who are present at the time of the birth. We want to explore the data to substantiate these beliefs. We evaluate the effect that missing data might have on analyses and inferences based on that data. Further, we perform multiple imputations for the missing data and then compare the inferences drawn from the incomplete dataset to those drawn from the imputed datasets. We also combine the baseline datasets (complete and completed) with the wave 2 dataset, to evaluate changes in marital status or relationship of parents, and the variables that might be significant predictors of these changes.

I. Introduction

As more and more children are being born outside marriage in the United States, policy makers and community leaders are becoming worried about the long-term effects. The Fragile Families and Child Wellbeing Study (FFCWS) is an attempt in this direction. It follows a cohort of 4898 births. The study was designed such that both parents were to be interviewed immediately after the baby is born and again in follow-up interviews when the child is one year old, three years old and five years old. In this paper, we look at the baseline dataset and note that about 22% of the birth fathers were not present to be interviewed, but all the mothers could be reached for the interviews. We wanted to see whether or not missing data could be a problem in any analyses that we may want to do. Also, because missing data is found to be an issue, we propose imputing for the missing fathers' data.

Section II describes the contents of the baseline dataset and tells us more about the purpose of the study. Section III outlines the methods used in handling missing data. In particular, we describe multiple imputation, the method we shall be using to adjust for the missing data. Section IV presents some results from the imputations and also comparisons between the incomplete dataset and imputed datasets. It also shows the results obtained after combining the baseline and wave 2 datasets. Section V discusses the conclusions of this paper and suggests more analyses that could be done.

II. The Fragile Families and Child Wellbeing Study Data

The Fragile Families and Child Wellbeing Study (FFCWS) is a joint effort by Princeton University's Center for Research on Child Wellbeing and Columbia University's Social Indicators Survey Center. This study is funded through a variety of sources including the National Institute of Child Health and Human Development, National Science Foundation, U.S. Department of Health and Human Services, California Healthcare Foundation, Ford Foundation and others.

The FFCWS follows a birth cohort of the children of (mostly) unwed parents over a five-year period. It was designed to provide new information on the relationships and capabilities of unwed parents, and how these affect their children. It also aims towards shedding light on the effects of policies on family formation and child wellbeing by addressing three areas of interest to policy makers and community leaders ─ non-marital childbearing, welfare reform and the role of fathers. As is stated in the Baseline National Report (McLanahan et al, 2003) of the study, "*by gaining a more complete* *understanding about the lives of unmarried parents, community leaders and policymakers can design programs that more effectively meet the needs of new, unmarried parents and thereby strengthen fragile families.*"

This study used a multi-stage stratified cluster sample. The sample of unwed parents is such that, when weighted, it is representative of all non-marital births to parents residing in U.S. cities with 200,000 or more people. The stratification was according to the policy environments and labor market conditions, rather than the geographical locations of the cities. The sampling was done in three steps: first the cities were sampled, then the hospitals were sampled within the cities and then births were sampled within the hospitals.

Baseline interviews were conducted in 75 hospitals in the following 20 cities across the United States:

Austin, TX	Detroit, MI
Newark, NJ	Pittsburgh, PA
New York City, NY	Richmond, VA
Boston, MA	Jacksonville, FL
Milwaukee, WI	Oakland, CA
San Jose, CA	Corpus Christi, TX
Baltimore, MD	Indianapolis, IN
Norfolk, VA	San Antonio, TX
Chicago, IL	Nashville, TN
Philadelphia, PA	Toledo, OH

The baseline interviews were conducted between February 1998 and September 2000. Each of the 4,898 births sampled and the corresponding 'family' was given a unique 'family identity number'. The baseline dataset contains 4,898 complete mother interviews (1,186 marital births and 3,712 non-marital births) and 3,830 complete father interviews. Thus, only 78% of the mothers have corresponding father interviews. The mothers' dataset contains information on 333 variables and the fathers' dataset on 338 variables. The variables contain demographic information (such as race, ethnicity, age and level of education) as well as information on mother and father relationship, marriage attitudes, health conditions, social support and family relationships, environmental factors, and awareness of government programs. The data also contains information on employment, income and economic well-being of both parents. The mothers' questionnaire design has the huge benefit that it asks mothers for data about the birth fathers, and this information can be used to help understand and impute for missing fathers.

As we have seen earlier, we have complete data on only 3,830 of the 4,898 fathers that were to be interviewed. This means that 1,068 fathers were missing. What do we do about the missing fathers? Does it matter to the study that 22% of the fathers could not be interviewed at the baseline? Are the fathers missing at random or are the missing fathers different from the ones interviewed? To begin to explore these questions, we looked at the following table of comparisons between the interviewed fathers and missing fathers, based on what the mothers said about them.

Table 1: Comparison of the Interviewed fathers versus the Missing fathers, based on what the mothers reported about them:

Mother reported ↓	**Interviewed fathers, n = 3,830**	**Missing fathers, n = 1,068**
1. Mother is married to Baby's father (BF)	28.09%	12.83%
2.BF gave money during pregnancy	64.36%	48.60%
3. BF suggested abortion	8.22%	16.10%
4. Race of BF		
White	29.24%	18.54%
Black/ African-American	48.90%	56.93%
American-Indian	4.13%	4.49%
5. Highest grade of school of BF		
Less than HS	28.64%	25.84%
HS	27.47%	29.03%
Post HS/ Some College	20.84%	15.54%
Bachelor's degree or more	11.07%	5.62%
6. Chance that mother will marry BF		
Pretty good/ almost certain	63.27% (n = 2,777)	29.36% (n = 906)
None/ very little	18.04% (n = 2,777)	54.30% (n = 906)

As can be seen in the Table 1, on at least some variables the missing fathers appear to be very different from the ones interviewed, based on what the mothers reported about the fathers. For example, of the fathers who were interviewed, 28.09% are reported (by the mothers) to be married to the mothers, whereas of the ones missing, only 12.83% are reported to be married to the mothers. Only 8.22% of the interviewed fathers are reported to have suggested abortion compared to 16.10% of the missing fathers. Another characteristic where the difference is striking is the reported race of the fathers. Among the interviewed fathers, there are reportedly 29.24% Whites and 48.90% Blacks or African-Americans, whereas among the ones missing there are 18.54% Whites and 56.93% Blacks or African Americans. Also, there are fewer fathers with some college (15.54%) and Bachelor's degree or more (5.62%) among the missing than among the ones interviewed (20.84% and 11.07%, respectively.) Notice that about 72% of the interviewed fathers are not married to the mothers, and for these cases, 63.27% of the mothers think that there is a pretty good or almost certain chance that they will get married to the baby's father. In the case of missing fathers, only 29.36% of the mothers think that there is a pretty good or almost certain chance of their marrying the baby's fathers, if they are not already married to them.

The variables of interest for the fathers, in our study are those of level of education, age, household relationship, race, ethnicity and income. Many studies show that these are the variables most likely to determine the future of the mother-father relationship and thus the well-being of their children. Therefore, it is important that we have complete information on these variables, and they will be the focus of our imputation efforts. The comparisons made in Table 1 give strong indication that the missing fathers were not missing at random. The missing fathers appear to be different from the ones interviewed. Taking these differences into account in our imputations should help to reduce bias from the differential missingness.

One possibility for imputation is that we could just use whatever the mother reported about the baby's father as his information. We will, in fact, use the mothers' reports as the starting point for our imputations, but we will use statistical models to help us improve over simply treating the mothers' reports as the "truth."

III. Handling Missing Data- Multiple Imputations

Imputation is a way to replace missing values by plausible values. Imputation lets us create completed datasets that can be used for future studies and that can be analyzed using standard statistical methods. There are many types of imputation, namely mean imputation, cold deck imputation, hot deck imputation, regression imputation and multiple imputation (Little and Rubin, 2002). Here we will use multiple imputations for handling the missing data (see Rubin, 1987).

Multiple imputation replaces each missing value with 2 or more plausible values, using a stochastic imputation procedure. We use multiple imputations instead of single imputation because multiple imputations allow us to account for the uncertainty due to imputation in later analyses. For datasets as large as ours, five imputations have proved to be enough. For more information on the theory and practice of multiple imputations, see Rubin, 1987.

We chose to impute for the following variables when they were missing for the baby's father: race, race-ethnicity[1], level of education, household relationship, income and age. Note that race, race-ethnicity, education, household relationship and income are categorical variables and age is a continuous variable.
(For the full questionnaires see http://opr.princeton.edu/archive/ff/).

Multiple imputations for Race, Race-ethnicity, Education, Age and Household Relationship

Imputations for these five variables were done differently from that for the income variable. For these variables, we had reports from both the mothers and the birth fathers in cases when the birth father responded to the survey. This helped us to compare the mothers' and interviewed fathers' answers on the same questions to see how well their answers matched. We used the 3,830 complete cases to find the conditional probabilities of the father being in a certain category, given what the mother reported. The following is the algorithm for the imputations:

1. Find the conditional probability of father being in a certain category, given what mother said.
2. Now for the 1,068 fathers who could not be interviewed:
   a. Generate a U (0,1) random number. Notice that this is a cumulative probability.
   b. Use the random number generated in (a.), the conditional probabilities from (1.) and what the mother reported about the missing father to impute what category the missing father might be in.
   c. Iterate (a.) and (b.) five times and record result each time.

[1] Race- ethnicity is a combination of race and ethnicity. The results in Section IV provide the categories.

When imputing for education, we had to be careful that the age and level of education were compatible. Thus, age was imputed first and then each time level of education was imputed, we checked if, for the imputed age, the imputed level of education was possible. For example, if the imputed age of the baby's father turned out to be 16, he could not possibly have a Bachelor's degree. In such cases, we re-imputed level of education. To be more specific, we set the following limits: a 16-year old could have less than high school level of education, a 17-year old could have less than high school, high school or post high school or some college, an 18 or 19-year old could have any of the previous or GED, a 20-year old could have any of the previous or refuse or skip or not know, and anyone 21 or older could have any of the previous or have a Bachelor's degree or more. In addition, the minimum acceptable age for the baby's fathers was set at 16 years, because that was age of the youngest fathers among those interviewed. Also, when imputing age for the missing fathers, we assumed that the mothers could not have been more than 10 years off the fathers' right ages, when reporting the fathers' ages.

Note that, for the variable Age, there were forty cases where the mother's report on father's age was missing. For these forty cases, instead of imputing from mother-reported father's age, we imputed based on the mother's age, such that the acceptable difference between mother's and father's ages was set to lie between -11 and 20, i.e. the mother could be between eleven years older and twenty years younger than the baby's father. We arrived at this range for the difference by plotting a histogram for the age differences, using the 3,830 complete cases.

Multiple imputations for Income

Imputing income for missing fathers was difficult because we had no information on the baby's fathers' income levels from the mothers. Thus, we needed a different method for imputation from that we used for the other variables. For imputing the income level of the baby's father, we used multi-category response logistic regression, with nine response categories[2]. Previous studies have shown that income level of baby's father could be dependent on, among other things, his age, race-ethnicity and level of education and household relationship as per him. We are interested in the household relationship as per father, because it is believed that his sense of responsibility towards his partner and children is affected by it and so is his income level. In addition, whether or not the father did regular work in the last week and whether or not he had a mental/physical condition that prevented him from regular work turned out to be important explanatory variables when we ran a stepwise logistic regression procedure. The interaction effects that turned out to be important were the age and race interaction, education and race interaction, and age, race and education interaction. We should note here that age is the only continuous variable among the explanatory variables; all the others are categorical variables. The income levels of the 40 (missing) fathers whose age was not reported by the mothers could not be imputed.

[2] For the categories see the results in Section IV.

Computing point estimates and standard errors

After we complete the five imputations for the desired variables, we have one incomplete dataset and five completed datasets. To compare results from the imputations, we first obtain the percentages in each of the categories, for each of the imputed variables, for each of the datasets separately.

For the age variable, we obtain the mean age and the standard error using methods for continuous variables. The algorithm for computing point estimates and standard error, for the other variables, is as follows:

1. $\hat{p}^* = \frac{1}{m}\sum_{i=1}^{m} \hat{p}_i^*$, is the point estimate from the m imputed datasets, where $\hat{p}_i^*$ is the point estimate for the i[th] imputed dataset[3].
2. $T = \bar{U} + \frac{m+1}{m} V$, is the variance estimate, where, $\bar{U} = \frac{1}{m}\sum_{i=1}^{m} \hat{Var}\left(\hat{p}_i^*\right)$, is the average of the estimated variances of $\hat{p}_i^*$, and

[3] $\hat{p}_i^*$ and $\hat{p}^*$ are proportions for the various categories.

$$V = \frac{1}{m-1}\sum_{i=1}^{m}\left(\hat{p}^*_i - \hat{p}^*\right)^2$$, is

estimated variance between the $\hat{p}^*_i$'s.

3. Thus, $\hat{p}^*$ and $\sqrt{T}$ give the point estimate of the proportion and standard error, respectively, for the various categories of each of the variables of interest, for the imputed datasets.

IV. Some Results

In this section, we present some results from the multiply imputed data sets and then compare them to the results we would get if we used the incomplete datasets. We will then be able to see whether or not imputation makes a difference to our analyses and inferences.

We use the following abbreviations:

CD= Complete dataset
CDD= Completed dataset

Table 2: Estimated average age of Baby's father(Standard Errors are given in parenthesis)

	CD n = 3,830	**CDD n = 4,898**
Mean age (S.E.), in years	27.942 (7.2688)	28.031 (7.3495)

Table 2 shows that the mean age of interviewed fathers is around 27.942 years, which is slightly less than 28.031 years, the mean age from the multiply-imputed datasets.

Table 3 shows that we are likely to underestimate the proportion of Blacks or African-Americans (48.83%) in the population under study, when we have incomplete data. After imputations, the estimated proportion of Blacks or African-Americans is 50.57%. Also, we tend to overestimate the percentage of Non-Hispanic Whites and Asians at 22.53% for the complete dataset. Once the missing fathers have been accounted for, this percentage goes down to 18.29%. Quite predictably, using the complete dataset leads us to overestimate the proportion of married (28.09%) and cohabiting (43.29%) people in this population. The imputed datasets show that there are an estimated 25% married couples and 39.41% cohabiters in the population of interest. The complete dataset underestimates the proportion of men with less than high school level of education (33.42%) and overestimates the proportion of men with a Bachelor's degree or more (11.04%) as compared to the imputed datasets (37.64% and 8.82%, respectively.) Thus, it appears that imputation does make a difference in the inferences that can be drawn from the datasets.

Table 3: Estimated probabilities of Baby's Father being in various categories (Standard Errors are given in parenthesis)

	CD, n= 3830	**CDD, n = 4898**
Race of BF		
White or Asian	*31.85% (0.753%)*	*29.96% (0.69%)*
Black	*48.83% (0.808%)*	*50.57% (0.76%)*
Am-Indian or other	*17.36% (0.612%)*	*17.74% (0.58%)*
Refuse, skip, etc.	*1.96% (0.224%)*	*2.05% (0.24%)*
Race-ethnicity of BF	**CD, n = 3830**	**CDD, n = 4898**
Hispanic	*27.73% (0.723%)*	*27.57% (0.65%)*
Non-Hispanic Whites and Asians	*22.53% (0.675%)*	*18.29% (0.56%)*
Non-Hispanic Blacks	*46.71% (0.806%)*	*48.26% (0.72%)*
Other	*3.03% (0.277%)*	*5.88% (0.35%)*
Household relationship as per BF		
Married	*28.09% (0.762%)*	*25.00% (0.63%)*
Cohabiting	*43.29% (0.801%)*	*39.41% (0.76%)*
Romantically involved, but not cohabiting	*15.72% (0.558%)*	*17.53% (0.67%)*
Other	*12.90% (0.542%)*	*18.06% (0.70%)*
Level of education of BF		
Less than High School	*33.42% (0.762%)*	*37.64% (0.75%)*
HS	*25.17% (0.701%)*	*23.90% (0.63%)*
GED	*7.18% (0.417%)*	*7.26% (0.40%)*
Post HS or some college	*22.98% (0.680%)*	*22.16% (0.62%)*
Bachelor's degree/more	*11.04% (0.506%)*	*8.82% (0.41%)*
Refuse, skip	*0.21% (0.074%)*	*0.22% (0.09%)*

Income of baby's father	**CD, n= 3830**	***CDD, n =4858***
Under $5K	*11.80% (0.52%)*	*13.37% (0.53%)*
$5K- $9,999	*11.36% (0.51%)*	*12.77% (0.52%)*
$10K- $14,999	*12.66% (0.54%)*	*13.27% (0.54%)*
$15K- $19,999	*10.89% (0.50%)*	*11.21% (0.46%)*
$20K- $24,999	*9.95% (0.48%)*	*10.08% (0.48%)*
$25K- $34,999	*11.98% (0.52%)*	*11.76% (0.50%)*
$35K and above	*18.21% (0.67%)*	*17.17% (0.56%)*

Some analyses to compare incomplete and imputed datasets

In the previous section we compared estimates for the original and multiply –imputed datasets on the variables that were imputed. Now we wish to explore the effects of the imputation on analyses using variables other than those imputed. First we explored how the mothers' views regarding the chances of getting married to the baby's fathers varied with respect to the different levels of education, income and race-ethnicity of baby's father, and present household relationship. Thus, we compared the complete and imputed datasets to see whether or not the proportions were different for the various categories of these variables.

Any mother who was not already married to the baby's father was asked what she thought the chance was that she would get married to the baby's father. The three categories for response were "No/ poor chance", "Fifty-fifty" and "Very good/ almost certain". The following tables show the cross tabulation of "almost certain/ pretty good chance" versus the different categories of the variables of interest.

We can see in Table 4 that the mothers report higher chances of getting married to the baby's father if he has High School, GED and Some College or Post High School levels of education, but they are still not as high as the estimates suggested by the complete dataset. Table 5 shows that mothers' marital expectations are highest when the fathers are Non-Hispanic Whites or Asians, but they are still not as high as suggested by the complete dataset, where we have not accounted for the missing fathers. As we see in Table 6, the mother-reported chance of getting married to baby's father increases as his income level increases and then falls for the two highest levels, but the estimated percentages are smaller for the completed dataset as compared to the complete dataset. Why the chances get smaller for the two highest income categories could be of some interest to sociologists.

Table 4: "Almost certain or Pretty Good chance of marrying the baby's father by Level of education of Baby's Father"- Point estimates with Standard errors in parenthesis

Level of education of BF	***CD, n = 2777***	***CDD, n = 3683***
Less than High School	*61.49% (1.47%)*	*52.62% (1.38%)*
HS	*63.32% (1.75%)*	*57.57% (1.71%)*
GED	*63.18% (3.12%)*	*57.08% (3.10%)*
Post HS or some college	*67.76% (1.94%)*	*55.97% (1.86%)*
Bachelor's degree or more	*57.61% (5.15%)*	*54.11% (5.04%)*

Table 5: "Almost certain or Pretty Good chance of marrying the baby's father by Race-ethnicity of Baby's Father"- Point estimates with Standard errors in parenthesis

Race-ethnicity of BF	***CD, n = 2777***	***CDD, n = 3683***
Hispanic	*69.20% (1.63%)*	*59.62% (1.54%)*
Non-Hisp. Whites/Asians	*79.88% (2.23%)*	*68.34% (2.36%)*
Non-Hisp. Blacks	*57.05% (1.27%)*	*50.35% (1.11%)*
Other	*58.27% (4.38%)*	*49.27% (4.43%)*

Table 6: "Almost certain or Pretty Good chance of marrying the baby's father by Level of income of Baby's Father"- Point estimates with Standard errors in parenthesis:

Income of baby's father	***CD, n = 2777***	***CDD, n = 3670***
Under $5K	*60.19% (2.38%)*	*49.59% (2.34%)*
$5K- $9,999	*61.11% (2.45%)*	*51.05% (2.38%)*
$10K- $14,999	*63.41% (2.38%)*	*54.38% (2.20%)*
$15K- $19,999	*71.69% (2.50%)*	*61.25% (3.08%)*
$20K- $24,999	*71.12% (2.72%)*	*61.32% (2.95%)*
$25K- $34,999	*66.67% (2.76%)*	*58.14% (2.74%)*
$35K / above	*70.49% (2.92%)*	*61.51% (2.96%)*

Next we combined the baseline data with that from wave 2, to obtain the change in relationship status of the parents. We created the "relationship progress" variable to check whether they had become more serious (for example, they were cohabiting in wave 1 and got married by wave 2), had remained in the same relationship or become less serious, by wave 2. We also created the "marry or not" variable, to check whether the couples who are not married to each other at the baseline get married by the second wave.

We ran a stepwise selection procedure on the complete dataset and the completed datasets, to predict whether or not the unmarrieds get married by the second wave. See Table 7 attached at the end. It turns out that the same variables are significant in predicting whether or not couples get married, the variables being, whether father has children with another woman, mother's race-ethnicity, whether mother trusts men, the kind of relationship the couple is in at the baseline, mother's marital expectations, her age, father's level of education and his level of income. Next, we ran a stepwise selection procedure to predict "relationship progress" from baseline to wave 2, for both the complete and completed datasets. See Table 8 attached at the end. For the complete data, the variables that turned out to be statistically significant are father's age, whether he was at a regular job last week, whether the couple have more than one child together, whether father has children with another woman, mother's race-ethnicity, her income level, her marital expectations, whether she trusts men and the kind of household relationship the couple is in, at the baseline. The completed datasets, on the other hand, did not select mother's level of income, whether she trusts men, or father's age, as significant variables. Instead, mother's level of education turned out to be statistically significant. This shows that it matters with what kind of partners the mothers are associated. Notice that, as soon as we account for the missing fathers, the predictors for "relationship progress", change.

The above analyses show that ignoring the fathers' missing data could lead to misleading inferences about the population of fragile families. If we used the incomplete dataset, we would be mostly overestimating the proportion of mothers who thought there was a very good or almost certain chance of getting married given the different characteristics of the baby's father. We expect that such differences between estimates based on the incomplete and multiply-imputed data will also arise in analyses involving other survey variables.

V. Conclusions and Future Work

We have seen that there are differences between fathers in the FFCWS who respond to the baseline survey and those who do not. We have used multiple-imputation procedures to create completed datasets on key demographic variables for the missing fathers. Multiple imputation helps to identify areas where we might make wrong predictions about the future of the parents' relationship if missing fathers are ignored. Also, the effects of ignoring the missing fathers are more visible in predicting "relationship progress" than in "marrying by the second wave." Our work thus far, however, is just the beginning. Some areas for future work include:

- Check what happens if we use only father's variables as predictors for both "relationship progress" and "marrying by wave 2."
- Study if and how the children's wellbeing (mental as well as physical) might be affected by the parents' relationship.

Acknowledgements:

I would like to thank The Ohio State University's Center for Survey Research for funding me in the summer of 2004, to do this research. Also I'd like to thank Dr. Elizabeth Stasny, (Department of Statistics, OSU) and Dr. Sharon Sassler, (Department of Sociology, OSU) for their invaluable advice and guidance.

References:

- CARLSON, M., MCLANAHAN, S. and ENGLAND P. (May 2004), "Union Formation in Fragile Families," Demography, Volume 41-Number 2; 237-261.
- LITTLE, R.J.A. and RUBIN, D.B. (2002), "Statistical Analysis with Missing Data," Hoboken, N.J.: Wiley-Interscience.
- MCLANAHAN S., GARFINKEL I., REICHMAN N., TEILTLER J., CARLSON M., AUDIGIER C.N., (March 2003), "The Fragile Families and Child Wellbeing Study- Baseline National Report," http://crcw.princeton.edu/fragilefamilies/files/nationalreport.pdf.
- MINCY, R.B. (1994), "Strengthening Fragile Families: A Proposed Strategy for the Ford Foundation Urban Poverty Program." New York: Ford Foundation.
- MINCY, R.B. and POUNCY H. (1997), "Paternalism, Child Support Enforcement, and Fragile Families," The New Paternalism: Supervisory Approaches to Poverty, Washington, DC: Brookings Institution.
- RUBIN, D.B. (1987), "Multiple Imputations for Nonresponse in Survey," New York: John Wiley.
- SASSLER, S.L. and MCNALLY J., (2003), "Cohabiting couples' economic circumstances and union transitions: a re-examination using multiple imputation techniques," Social Science Research 32 (2003) 553-578.

- SASSLER, S.L., ROY, S. and STASNY, E.A. (2005), "Women's Marital Expectations and Subsequent Union Formation in Fragile Families." (Submitted to *Demography*.)

Table 7: Will the unmarried be married by wave 2? Coefficients from a Multinomial Regression model (reference group= unmarried at wave 2)

** p ≤ 0.05; * p ≤ 0.10

Variables	CD Est.	CD S.E.		CDD Est.	CDD S.E.	
Intercept	3.82	0.50	**	3.52	0.48	**
Mother reported						
BF's children (Ref= other)						
Kids with another woman	0.06	0.22		0.12	0.22	
Doesn't	0.48	0.21	**	0.63	0.21	**
Mother's race-ethnicity(Ref= NH Af-Am)						
Hispanic	0.33	0.16	**	0.36	0.15	**
NHWhite	0.19	0.17		0.23	0.16	
Other	0.10	0.37		0.31	0.37	
Mother doesn't trust men	-0.25	0.11	**	-0.21	0.10	**
Household relationship (Ref=Cohabiting)						
Non-romantic	-0.32	0.22		-0.47	0.21	**
Romantically involved	-0.04	0.15		0.00	0.14	
Marital expectation(Ref= Poor chance)						
Good chance	0.89	0.17	**	0.82	0.15	**
Fifty-fifty	0.17	0.19		0.10	0.17	
Mother's age	0.03	0.01	**	0.03	0.01	**
Father's attributes						
Level of education (Ref= LTHS)						
Do not know	1.07	1.00		0.84	0.97	
HS_GED	-0.39	0.28	**	-0.30	0.27	*
Some College	-0.17	0.28		-0.10	0.27	
BA or more	0.22	0.35		0.30	0.34	
Income level (Ref= <$5K)						
Do not know	-0.08	0.20		-0.01	0.20	
$5K-$9999	-0.29	0.19		-0.33	0.20	*
$10K-$14999	0.11	0.16		0.08	0.17	
$15K-$19999	0.26	0.17		0.24	0.16	
$20K-$24999	-0.05	0.19		-0.03	0.20	
$25K-$34999	0.47	0.16	**	0.45	0.16	**
$35K or more	-0.05	0.20		-0.03	0.19	

Table 8: Predicting relationship progress from baseline to wave 2: Coefficients from a Multinomial Regression model (Reference group is "Less serious at wave 2)

** p ≤ 0.05; * p ≤ 0.10

Variables	CD More serious		Same		CDD More Serious		Same	
Intercept	2.46	**	1.59	**	2.33	**	1.11	**
BF's age	0.03	**	0.03	**				
BF was at a regular job last week	0.35	**	0.2		0.35	**	0.19	*
More than 1 child together	0.28	**	0.08		0.42	**	0.13	
BF's kids(Ref=BF doesn't have kids with another woman)								
BF has kids with another woman	-0.58	**	-0.4	**	-0.64	**	-0.32	**
other	3.2	**	2.17	**	3.2	**	1.82	**
Mother's race-ethnicity(Ref= NH Af-Am)								
Hispanic	1.05	**	0.86	**	0.92	**	0.58	**
NH White	0.67	**	0.42	**	0.72	**	0.49	**
Other	0.5		0.05		0.42		0.26	
Mother's Income(Ref=<$5K)								
Don't know	-0.1		-0.11					
$5K-$9999	0.67	**	0.51	**				
$10K-$14999	-0.04		0.02					
$15K-$19999	0.12		-0.01					
$20K-$24999	0.07		-0.25					
$25K and more	0.10		-0.21					
Mother doesn't trust men	0.4	**	0.27	**				
Marital Expectation(Ref= Poor chance)								
Good chance	1.91	**	0.61	**	1.97	**	0.7	**
Fifty-fifty	1.32	**	0.37	**	1.35	**	0.3	
Household relationship(Ref= Romantic)								
Cohabiting	1.23	**	0.76	**	1.11	**	0.79	**
Non romantic	16.61	n/a	16.22	n/a	15.96	n/a	16.63	n/a
Mother's level of education(Ref= LTHS)								
HS or GED					0.57	**	0.13	
Some College					0.44	**	0.56	**
BA or more					0.69	**	0.03	