APPENDIX A: DESCRIPTION OF EARNINGS ESTIMATIONS SAMPLES

We begin with the full sample of fathers from the Fragile Families and Child Wellbeing Study (*N* = 4,898). Then, we apply the following inclusion/exclusion criteria across all four earnings estimation samples (Simulations 1, 2, 3, and 4–6):

- Include only fathers who are unwed at the baseline interview (*N* = 3,710).
- Exclude fathers whose children are deceased (21 observations) at the one-year interview (*N* = 3,689).
- Exclude fathers whose children are adopted (14 observations) at the one-year interview (*N* = 3,675).
- Include only fathers who are nonresident at the one-year interview (*N* = 2,063).

In our study sample (*N* = 2,063), 853 fathers reported their earnings at the one-year interview. Our goal is to impute fathers' earnings for the full sample of fathers (*N* = 2,063), according to six simulations.

Simulation 1: Assortative Mating/Single Imputation

The estimation model uses standard earnings equation variables based on the following assortative mating criteria applied to mothers' demographic characteristics:

- Father's race = mother's race.
- Father's age = mother's age + two years.
- Father's education = mother's education.

Simulation 2: Actual Mothers' Reports of Fathers' Characteristics/Single Imputation

The estimation model uses standard earnings equation variables based on mothers' reports of fathers' demographic characteristics.

Simulation 3: Fathers' Self-reports Plus Previously Unobserved Characteristics/Multiple Imputation

The expanded estimation model uses standard earnings equation variables based on fathers' self-reports, plus a wide range of fathers' self-reported characteristics, mothers' self-reported characteristics, and mothers' reports of fathers' characteristics.

Simulations 4–6: Ineligible Fathers

Estimation model is similar to Simulation 3. Additionally, this model predicts zero earnings for ineligible fathers (202 observations):

- Deceased = 17
- Negative DNA test = 14
- Unknown = 58
- Not told of pregnancy = 69
- Denies paternity = 37
- Other ineligibility = 7

Table A1. Single Imputation Estimation Models for Fathers' Earnings

Variable	Assortative Mating Demographics	Actual Mothers' Reports of Fathers' Demographics
Race		
White (ref.)		
Black	–3,636*	–4,266*
	(1,704)	(1,799)
Hispanic	–496	–283
	(2,090)	(2,126)
Age		
Younger than 21 years (ref.)		
21–29 years	2,264	2,273
	(1,849)	(1,576)
30 years or older	7,761**	7,869**
	(2,231)	(1,850)
Education		
Less than high school (ref.)		
High school	1,295	4,223**
	(1,504)	(1,439)
Some college	3,658*	8,735**
	(1,634)	(1,738)
College degree	13,228**	17,251**
	(3,412)	(3,653)
Constant	18,296**	16,028**
	(–2,196)	(–2,189)
R^2	0.05	0.10

Note: Standard errors are in parentheses.

$^{*}p \leq .05$; $^{**}p \leq .01$

APPENDIX B. TECHNICAL APPENDIX: MULTIPLE IMPUTATION STRATEGY

Most studies that use survey or administrative data have to decide how to deal with missing data. In this article, we use multiple imputation (MI) because it has several advantages over conventional strategies to handle missing data (e.g., complete cases, complete variables, nonresponse weighting). MI yields improved standard errors because it accounts for uncertainty about the model and the sample. It also relies on more plausible assumptions about the missing data mechanism. MI posits that individuals who have the same values on all observed covariates should be expected to have the same probability of having missing data. This is known as missing-at-random (MAR). MAR is a weaker (more plausible) assumption than that underlying complete cases, which posits that individuals with missing data are a completely random subset of the full sample. This assumption is known as missing completely at random (MCAR), a stricter assumption than MAR. For a comparison using simulation models, see Hill (2004), Rubin (1987), and Schafer (1997).

Our extended MI earnings model captures the relative disadvantage of nonrespondent fathers because it includes a wide range of variables based on fathers' self-reports, mothers' self-reports, and mothers' reports of fathers, which together are predictive of nonresponse and lower earnings. Beyond the larger array of predictor variables in the model, the predictor variables are themselves multiply imputed, and thus these distributions now more accurately reflect the relative disadvantage of the full distribution of fathers.

The basic concepts underlying our multiple imputation strategy are as follows (Allison 2001):

1. Impute the missing values using an appropriate model that incorporates random variation. The 88 variables in our model were derived from three sources: (1) fathers' self-reports; (2) mothers' self-reports; and (3) mothers' reports of fathers.
2. Do this *M* times, producing *M* "complete" data sets. In our study, *M* equals 5; that is, five complete-case data sets were imputed.
3. Perform the desired analysis on each data set using standard complete-data methods.
4. Average the values of the parameter estimates across the *M* samples to produce a single-point estimate.
5. Calculate the standard errors by (a) averaging the squared standard errors of the *M* estimates, (b) calculating the variance of the *M* parameter estimates across samples, and (c) combining the two quantities using a simple formula. Standard errors that account for inter- and intra-dataset variation are computed according to the rules laid out by Rubin (1987).

The multiple imputation computations are implemented in Stata using the multivariate imputation by chained equations (MICE) method of multiple multivariate imputation described by van Buuren, Boshuizen, and Knook (1999). The application in Stata was developed by Patrick Royston (2004).

Table B1 reports the overall response rates for fathers at each of the first two waves of data. The baseline response rate for all fathers is 78%. At the one-year interview, it is about 10 percentage points lower than at baseline. This pattern is similar for both married and unmarried fathers, although response rates are consistently higher for married than unmarried fathers—about 89% and 75%, respectively, at baseline compared with 81% and 65%, respectively, at the one-year interview.

Table B1. Percentage of Fathers Who Were Interviewed

Sample	Baseline	One-Year Interview
Full Sample	78	69
Married	89	81
Unmarried	75	65

The vast majority of the missing data we impute concern fathers who were not interviewed and information about fathers that we could not accurately ascertain from mothers. Of the 88 variables in the MI model, roughly one-fourth to one-third of the data are imputed across all father-reported variables. By contrast, similar proportions of missing data are found for only one-third of the mother-reported variables. For the remaining two-thirds of the mother-reported variables, only 10% of the data are missing. The number of missing variables per observation ranges from a high of 80% for one observation to complete data for 15% of the 4,898 observations. Fully half of the observations have eight or fewer missing variables, while for the 10% of fathers with the most missing data, one-third to one-half of the variables in the model are imputed. Table B2 provides a comparison of the distribution of selected baseline characteristics of fathers using complete-case data and multiple imputation data. For a more detailed comparison of imputation methods (complete cases, single imputation, and multiple imputation) that account for missing data in the Fragile Families and Child Wellbeing Study, see Sinkewicz (2006).

Table B2. Selected Baseline Characteristics of Fathers: Comparison of Complete Cases (CC) and Multiple Imputation (MI)

	Full Sample (N = 4,898)		Married (N = 1,188)		Unmarried (N = 3,710)	
Variable	CC	MI	CC	MI	CC	MI
Race						
White	34.1	32.8	43.0	41.5	11.8	12.2
Black	26.8	28.4	25.5	26.1	55.7	56.0
Hispanic	32.7	32.2	24.8	25.1	29.0	27.0
Other	6.4	6.6	6.8	7.3	3.5	4.9
Age						
<21 years	9.7	9.8	2.0	2.2	18.1	17.2
21–29 years	42.6	41.9	36.1	34.5	54.7	51.9
30+ years	47.7	48.3	61.9	63.3	27.1	31.0
Agree About Relationship	99.2	98.4	98.7	97.9	98.6	97.4
Other Kids						
0	39.1	37.2	36.4	34.7	44.5	39.3
1, 2, or 3	57.0	58.1	59.7	60.5	49.6	53.9
4 or more	4.0	4.7	3.9	4.7	5.9	6.9
U.S.-Born	78.5	77.9	72.6	71.0	85.2	83.2
Education						
Less than high school	27.1	27.6	17.1	17.0	39.7	38.8
High school	28.6	27.6	23.1	22.0	35.9	33.9
Some college	25.4	24.1	28.3	26.4	21.0	19.8
Bachelors degree or more	18.9	20.7	31.5	34.6	3.3	7.5
Worked Last Week	86.8	86.2	92.7	92.7	75.9	74.7
In School Last Week	0.6	19.1	0.7	14.9	1.0	31.8

(continued)

(Table B2, continued)

	Full Sample (N = 4,898)		Married (N = 1,188)		Unmarried (N = 3,710)	
Variable	CC	MI	CC	MI	CC	MI
Annual Regular Earnings						
$0	3.1	5.3	2.2	2.3	5.7	10.7
$1–$9,999	15.5	15.5	7.0	6.9	32.6	28.8
$10,000–$19,999	23.8	23.3	17.0	16.5	29.3	27.9
$20,000–$29,999	26.2	24.9	27.6	25.9	22.7	21.1
$30,000–$49,999	13.8	15.0	18.4	21.0	6.5	8.8
$50,000 or more	17.7	16.0	27.7	27.5	3.3	2.7
Annual Informal Earnings						
$0	74.3	66.8	79.5	71.0	70.6	60.3
$1–$9,999	22.5	29.8	17.5	25.8	26.4	36.1
$10,000–$19,999	1.3	1.8	1.4	1.9	1.4	2.4
$20,000–$29,999	1.7	1.4	1.6	1.3	1.2	0.8
$30,000–$49,999	0.2	0.2	0.0	0.0	0.3	0.2
$50,000 or more	0.1	0.0	0.0	0.0	0.1	0.1
Illegal Work Activity	1.8	2.2	0.3	0.3	4.1	4.2
Training: Government/ Vocational/Military	41.8	39.7	41.5	39.7	41.7	38.1
Military	14.5	17.7	15.8	18.7	10.5	16.3
Car Ownership	78.6	74.9	87.6	85.5	56.1	50.0
Bad Health	7.7	9.3	6.0	6.4	8.2	11.8
Incarceration	2.9	3.9	0.7	1.6	7.3	8.3
Religiosity	35.9	35.5	46.3	47.4	25.5	24.7

Note: The full sample is weighted to be representative of all births in U.S. cities with a population greater than 200,000.

REFERENCES

Allison, P. 2001. *Missing Data*. New York: Sage.

Hill, J. 2004. "Reducing Bias in Treatment Effect Estimation in Observational Studies Suffering From Missing Data." Working Paper 04-01. ISERP, Columbia University.

Royston, P. 2004. "Multiple Imputation of Missing Values." *The Stata Journal* 4(3):227–41.

Rubin, D. 1987. *Multiple Imputation for Non-response in Surveys*. New York: John Wiley and Sons.

Schafer, J. 1997. *Analysis of Incomplete Multivariate Data*. London: Chapman and Hall.

Sinkewicz, M. 2006. "The Mental Health of Men: Profile and Life Trajectories of Urban American Fathers." Doctoral dissertation. School of Social Work, Columbia University.

van Buuren, S., H. Boshuizen, and D. Knook. 1999. "Multiple Imputation of Missing Blood Pressure Covariates in Survival Analysis." *Statistics in Medicine* 18:681–94.